DIGESTIF: A universal quality standard for the control of bottom-up proteomics experiments

Lebert, Dorothée ; Louwagie, Mathilde ; Goetze, Sandra ; Picard, Guillaume ; Ossola, Reto ; Duquesne, Caroline ; Basler, Konrad ; Ferro, Myriam ; Rinner, Oliver ; Aebersold, Ruedi ; Garin, Jérôme ; Mouz, Nicolas ; Brunner, Erich ; Brun, Virginie

Abstract: In bottom-up mass spectrometry-based proteomics analyses, variability at any step of the process, particularly during sample proteolysis, directly affects the sensitivity, accuracy, and precision of peptide detection and quantification. Currently, no generic internal standards are available to control the quality of sample processing steps. This makes it difficult to assess the comparability of MS proteomic data obtained under different experimental conditions. Here, we describe the design, synthesis, and validation of a universal protein standard, called DIGESTIF, that can be added to any biological sample. The DIGESTIF standard consists of a soluble recombinant protein scaffold to which a set of 11 artificial peptides (iRT peptides) with good ionization properties has been incorporated. In the protein scaffold, the amino acids flanking iRT peptide cleavage sites were selected either to favor or hinder protease cleavage. After sample processing, the retention time and relative intensity pattern of the released iRT peptides can be used to assess the quality of sample workup, the extent of digestion, and the performance of the LC-MS system. Thus, DIGESTIF can be used to standardize a broad spectrum of applications, ranging from simple replicate measurements to large-scale biomarker screening in biomedical applications.

DOI: https://doi.org/10.1021/pr500834z

DIGESTIF, a universal quality standard for the control of bottom-up proteomics experiments

Dorothée Lebert[1]§, Mathilde Louwagie[2,3,4]§, Sandra Goetze[5,6]§, Guillaume Picard[1], Reto Ossola[7], Caroline Duquesne[1], Konrad Basler[5], Myriam Ferro[2,3.4], Oliver Rinner[7], Ruedi Aebersold[6,8], Jérôme Garin[2,3,4], Nicolas Mouz[1], Erich Brunner[5]*, Virginie Brun[2,3.4]*

[1] Promise Advanced Proteomics, Grenoble, F-38040, France

[2] Univ. Grenoble Alpes, iRTSV-BGE, F-38000 Grenoble, France

[3] CEA, iRTSV-BGE, F-38000 Grenoble, France

[4] INSERM, BGE, F-38000 Grenoble, France

[5] Institute of Molecular Life Sciences (IMLS), University of Zurich, Zurich, Switzerland

[6] Department of Biology, Institute of Molecular Systems Biology, ETH Zurich, CH-8093 Zurich, Switzerland

[7] Biognosys AG, Zurich, Switzerland

[8] Faculty of Science, University of Zurich, Zurich, Switzerland

§ These authors contributed equally to this work

* Corresponding authors:

Erich Brunner

Tel: +41 44 635 31 59

Fax: +41 44 635 68 64

erich.brunner@imls.uzh.ch

Virginie Brun

Tel: +33 4 38 78 96 57

Fax: +33 4 38 78 50 51

virginie.brun@cea.fr

ABSTRACT

In bottom-up mass spectrometry-based proteomics analyses, variability at any step of the process, particularly during sample proteolysis, directly affects the sensitivity, accuracy and precision of peptide detection and quantification. Currently, no generic internal standards are available to control the quality of sample processing steps. This makes it difficult to assess the comparability of MS proteomic data obtained in different experimental conditions. Here, we describe the design, synthesis and validation of a universal protein standard, called *DIGESTIF,* that can be added to any biological sample. The *DIGESTIF* standard consists of a soluble recombinant protein scaffold to which a set of 11 artificial peptides (iRT peptides) with good ionization properties has been incorporated. In the protein scaffold, the amino acids flanking iRT peptide cleavage sites were selected to either favor or hinder protease cleavage. After sample processing, the retention time and relative intensity pattern of the released iRT peptides can be used to assess the quality of sample workup, the extent of digestion and the performance of the LC-MS system. Thus, *DIGESTIF* can be used to standardize a broad spectrum of applications, ranging from simple replicate measurements up to large-scale biomarker screenings in biomedical applications.

KEYWORDS: Proteolysis, digestion, standard, proteomics, mass spectrometry, biomarkers.

ABBREVIATIONS

CE: Collision Energy

iRT peptide: indexed Retention Time peptide

PSAQ: Protein Standard Absolute Quantification

QconCAT: Quantification Concatemer

RT: Retention Time

SRM: Selected Reaction Monitoring, also referred to as MRM (Multiple Reaction Monitoring)

INTRODUCTION

Over the last two decades, extensive technological and methodological advances have led to the widespread use of mass spectrometry (MS)-based bottom-up proteomics. MS-based analysis is the method of choice for large-scale exploration of proteomes in biological systems and has significantly contributed to the comprehension of a variety of biological processes [1]. In the field of medical diagnostics, quantitative MS has been an integral part of numerous biomarker discovery and evaluation studies [2,3]. The rapid growth of therapeutic classes employing protein-based compounds, in particular monoclonal antibodies, has led to the implementation of MS at all stages in their development for structural and analytical characterization [4].

Bottom-up proteomics - which involves the digestion of proteins into peptides followed by identification and quantification of the resulting peptides by LC-MS - remains the method of choice for proteome characterization and quantification [5,6] in spite of impressive advances in top-down proteomics [7] and affinity-based proteomic measurements [8]. Because of this, elaborate multistep protocols have been established to support the application of bottom-up proteomics even for complex samples. These protocols typically include the extraction of proteins from tissues, cells or body fluids; proteins are then cleaved and the resulting peptide mixtures are separated on a reversed-phase column prior to MS analysis. To achieve high quality and highly reproducible results, all steps in this complex procedure must be carefully validated and controlled. To date, attempts to standardize the complex workflow have been moderately

successful [9] mainly because of the stochastic steps involved [10-13]. Because of these steps, quality control and assessment efforts in bottom-up proteomics have tended to focus on the use of various reference standards to check performance at specific stages in the process. Thus, specific studies have been conducted with different types of standards to monitor LC separation, to optimize mass spectrometer performance [14], to assess protein digestion [15,16] or to evaluate the reproducibility of measurements [12,17,18].

In an attempt to expand on these, Eyers and coworkers [19] presented QCAL, a recombinant protein standard consisting of a set of concatenated peptides, in 2008. After trypsin digestion, QCAL provides a stoichiometrically controlled peptide mixture allowing the concomitant assessment and optimization of multiple MS-instrument parameters on a wide variety of instrument platforms.

More recently, Escher and coworkers [20] described a method to standardize chromatographic retention time as an iRT (indexed retention time) using a set of standard peptides (iRT peptides). These peptides are now commercially available as a kit. The iRT method was designed to precisely identify and correct for variations in retention time across a series of LC-MS measurements. This is particularly relevant for LC-SRM analyses. The iRT peptide set consists of 11 artificial peptide standards of varying hydrophobicity; these peptides are eluted at regular intervals from classical C18 reverse-phase columns using standard LC gradients. When spiked into the biological matrix to be studied, they serve as standards and make it possible to accurately predict the retention time of peptides monitored in LC-MS experiments. By monitoring

peptide precursors and defined SRM transitions, the iRT peptide signals also give indications on the performance of the MS system and provide a backdrop for statistical target validation and reproducible protein quantification in SRM assays, using the mProphet or Skyline softwares [21,22]. While undoubtedly a very useful tool, the iRT peptides are spiked into samples post-digestion, they therefore cannot be used to monitor sample preparation variance.

These sample processing steps are crucial, particularly for quantitative proteomics applications in clinical settings. It is therefore necessary to be able to rigorously evaluate and benchmark them. To meet this need, Percy and colleagues [23] recently described two sets of reagents to specifically help with standardization of biomarker detection in human plasma samples. The first set ("LC-MS test kit") is designed to check LC-MS performance. It contains plasma digests spiked with labeled peptides representing 43 high-to-moderate abundance plasma proteins. The second set ("QC test kit") is used to evaluate the entire analytical workflow (including proteolysis). It is composed of plasma samples, trypsin and the mix of 43 surrogate labeled peptides. Using these two kits, the reproducibility and reliability of quantitative biomarker analyses across laboratories can be enhanced. They also pave the way for benchmarking proteomics-based biomarker detection in human plasma for diagnostic applications. However, it is important to note that the reagents contained in these kits have been specifically optimized for use with plasma, which is the most complex and widely used human protein sample. Thus, they may not be suitable for use with other sample types, and it might be necessary, for each sample

and/or proteomics workflow, to compile a distinct set of test- and quality control reagents.

Although these solutions have significantly improved standardization of bottom-up proteomics experiments, the major source of variation affecting protein sequence coverage, detection sensitivity and, in particular, protein quantification remains proteolysis [16,24-31]. In bottom-up proteomics experiments, the most commonly used protease is trypsin. The rules for trypsin specificity (also referred to as “Keil rules”) were initially defined in 1992 as "C-terminal cleavage to arginine and lysine but not, or at very low-frequency, before proline" [32]. These rules were experimentally validated in 2004 on a large LC-MS/MS data set corresponding to the mouse liver proteome [33]. Then, using 14.5 million LC-MS/MS spectra from *Shewanella oeidensis,* Rodriguez and co-workers [34] further refined these rules. These authors identified amino acids flanking cleavage sites and strongly influencing digestion efficiency. These data were further confirmed by subsequent studies using different datasets [28,35].

In addition to the effects of surrounding sequences and protein structure, proteolytic cleavage of target proteins may be altered in a variety of body fluids or tissue samples. To assess this, various monitoring concepts have been developed. In 2005, Cutillas and coworkers [36] described the use of fetuin as an internal standard to control in-gel digestion. Their results were promising, but unfortunately this standard was not assessed in other proteolysis conditions. Additionally, as fetuin is a plasma protein, it is not suited to the study of plasma or serum samples. In 2012, Burkhart and coworkers proposed the use of a

monolithic column HPLC setup with UV detector to evaluate digestion efficiency and reproducibility prior to LC-MS/MS analysis[16]. An alternative approach is to spike samples with isotopically labeled standards before protein digestion. The best-known examples of these are Quantification Concatamers (QconCATs) and Protein Standards for Absolute Quantification (PSAQ™ standards) [37,38]. Both of these standards have mainly been used for the absolute quantification of specific target proteins in a mixture, based on the assumption that the external reference molecule and the endogenous counterpart share a similar fate during proteolysis and sample work-up. This assumption is more likely to be true for PSAQ standards than for QconCATs, as the peptides released from the QconCAT are not in the same sequence context as peptides released from the endogenous proteins [39]. This can be problematic as complex protein mixtures not only consist of proteins that are easily cleaved, but also contain highly structured species that may, in part, resist efficient proteolysis. QconCATs (including QCAL) are not predicted to form significant regions of secondary structure. In fact, they are designed and optimized for the most efficient release of their surrogate peptides [24,39]. While this may be appropriate for the specifically targeted proteins, it means they are poorly suited for evaluation of the digestion efficiency in complex mixtures. Similarly, it would not be appropriate to extrapolate the digestion performance from PSAQ standards to the whole proteome because they also reflect - at best - the properties of a subset of proteins in a complex sample.

Thus, the proteomics community needs a single-protein standard that (i) reflects the digestion properties of a complex protein mixture and (ii),

simultaneously allows detailed assessment of multiple sample analysis parameters (digestion conditions and rate of proteolysis, LC-MS performance). This is especially required in biomarker development studies that involve the digestion of highly complex matrices (serum/plasma) and quantify selected proteins across multiple samples. To meet this need, we have generated and validated a new generic standard, *DIGESTIF,* that has been specifically designed to release indicator peptides directly reflecting how well a generic protein sample is digested. In this article, we describe how *DIGESTIF* can be used to select optimal digestion conditions and assess proteolysis efficiency in bottom-up proteomics experiments to improve the reliability of MS-based biomarker studies.

EXPERIMENTAL SECTION

DIGESTIF molecular modeling and cloning

Using the PyMOL Molecular Graphics System software (www.pymol.org), the optimal *DIGESTIF* protein was designed by molecular modeling of three different variants, considering different sequential arrangements of the iRT peptides in the PBP2x scaffold [40]. For the variant displaying the most favorable design in terms of theoretical protein solubility and stability, a synthetic gene sequence was determined and optimized for *E. coli* codon usage. This sequence was purchased from GeneArt/Life Technologies (Regensburg, Germany). After digestion with NdeI and XhoI restriction enzymes (NEB, Evry, France), the synthetic gene was inserted into the PET33b vector (Novagen/Merck Millipore, Molsheim, France). This vector adds an N-terminal hexahistidine tag to the target protein and provides kanamycin resistance. The plasmid was chemically transformed into *E. coli* DH5α cells (NEB) for amplification. After purification, the plasmid DNA sequence was verified (LGC Genomics, Berlin, Germany). The pET33b-*DIGESTIF* vector was then transformed into a BL21(DE3) *E. coli* strain (Novagen/Merck Millipore), which is auxotrophic for lysine and arginine (genotype: *lysA-, argA-*). This strain [41] was specifically developed for the expression of proteins labeled with different isotopic forms of lysine and arginine.

DIGESTIF expression and isotope-labeling

Transformed auxotrophic *E.coli* cells were selected on plates containing kanamycin-supplemented LB medium. A positive clone was selected and grown in 1 L of a specific medium optimized for isotope labeling. This medium contained: M9 minimum salts supplemented with 2 mM $MgSO_4$ and 100 µM $CaCl_2$, 0.4% glucose, 50 µg/ml kanamycin, and 100 mg/L each [$^{13}C_6$, $^{15}N_4$] arginine and [$^{13}C_6$, $^{15}N_2$] lysine (Eurisotop, Saint Aubin, France). When the OD reached 0.5, expression of the *DIGESTIF* protein was induced by adding 1 mM IPTG to the media. Cell culture was maintained for 4 h at 37 °C.

DIGESTIF solubilization, purification and refolding

E. coli cells were harvested by centrifuging for 20 min at 5000 g and 4 °C. The pellet was suspended in 50 mL of Bugbuster (Novagen/Merck Millipore) supplemented with lysosyme at 200 µg/mL (Euromedex, Souffelweyersheim, France), 1 tablet of Complete™ EDTA Free (Roche Diagnostics, Meylan, France) and benzonase at 25 units/µl (Novagen/Merck Millipore). After 30 min at room temperature, the cell extract was centrifuged for 30 min at 15 000 g and 4 °C. The *DIGESTIF* protein was present in inclusion bodies. These were solubilized in a buffer containing 8 M urea, 20 mM Tris pH 7.5, 500 mM NaCl and 20 mM imidazole. The solubilized His-tagged *DIGESTIF* protein was purified using IMAC on Ni-Sepharose 6 Fast flow resin (GE Healthcare, Velizy, France) in accordance with the manufacturer's recommendations. Protein was eluted from the purification resin by applying an imidazole gradient. To refold and solubilize the

eluted *DIGESTIF* protein, the purification mixture was submitted to successive dialyses in buffers containing 20 mM Tris, 150 mM NaCl and decreasing urea concentrations (8/6/4/2/1/0.5/0.25/0 M urea). The purity of the resulting *DIGESTIF* standard was checked by SDS-PAGE analysis and sensitive Imperial Protein Stain staining (Thermo Fisher Scientific, Villebon sur Yvette, France). Based on this, purity was estimated to be greater than 90% (Supplementary Figure S1). *DIGESTIF* was then quantified using a microBCA assay (Sigma Aldrich).

DIGESTIF **proteolysis and peptide oxidation**

For LC-MS/MS analysis and to determine sequence coverage, the *DIGESTIF* standard was digested overnight at 37 °C in-solution using a 1/100 (w/w) trypsin-to-protein ratio (sequencing grade Trypsin, Promega, Charbonnières, France). To assess digestion kinetics, two different trypsin-to-*DIGESTIF* ratios were used: 1/100 (w/w) and 1/1000 (w/w). In these assays, digestion was also performed at 37 °C, but incubation time varied from 2 min to 24h, and digestion was stopped by adding formic acid (0.4% final concentration). After digestion, peptides were desalted with ZipTip C18 Pipette Tips (Millipore) before drying by vacuum centrifugation. In-solution oxidation was performed by adding 20 µL of 7% H_2O_2 / 0.5% formic acid to the sample. After at least 20 minutes' incubation at room temperature, peptides were injected onto the LC system.

Serum proteolysis experiments

Human male AB serum was obtained from Sigma Aldrich. Serum aliquots (6 µl, ~300 µg of serum proteins) were spiked with 1 µg (12.1 pmoles) of *DIGESTIF* (1/300, w/w) and digested into peptides using different protocols presented in Table 1. For these experiments, trypsin (sequencing grade) and trypsin/Lys-C mix (proprietary composition) were both purchased from Promega. Rapigest™ was obtained from Waters. In some experiments, samples were reduced using TCEP (2 mM, 30 min in the dark) and alkylated using iodoacetamide (10 mM, 30 min) prior to digestion (Figure 4/Test 10, Figure 7/Test 10). After proteolysis, each sample was spiked with unlabeled *DIGESTIF* peptides which served as quantification standards (12.1 pmoles of each peptide, AQUA QuantPro™ peptides, Thermo Fisher Scientific). Samples were desalted on C18 Macro SpinColumns (Harvard apparatus, Les Ulis, France). Samples that had not been reduced and alkylated were oxidized after proteolysis, as described above.

Experiments in human urine and mouse liver lysate

A urine sample (1 mL) was obtained from a healthy donor, member of a clinical research cohort established for urinary biomarker investigation. The protocol for this clinical research was approved by the local hospital's institutional review board, and the donor provided written informed consent for participation in this research. Mouse liver (C57BL/6 strain) was collected and provided by an

animal research facility operating in line with French and European guidelines and regulations for animal testing. Liver lysate was prepared by homogenizing frozen tissue at 4 °C in 3 volumes of lysis buffer (250 mM sucrose, Tris 50 mM pH 7.4). The homogenate was centrifuged for 10 min at 12 000 g and 4 °C and the supernatant was collected before protein quantification using a micro BCA assay. *DIGESTIF* standard (1 µg or 12.1 pmoles) was spiked into 8 µl (equivalent 300 µg total protein) of mouse liver lysate or 1 mL (equivalent 300 µg total protein) of urine, and different proteolysis protocols were assessed (see Table 1 and Figure 5). For urine samples, Filter aided sample preparation (FASP) using 10 MWCO spin filters (Merck Millipore, Fontenay sous Bois, France) was performed as previously described [42,43]. Slight variations on the FASP protocol (proteolysis with trypsin or double digestion with trypsin/Lys-C mix, absence or presence of reduction/alkylation treatment) were applied as detailed in Table 1 and Figure 5 (Tests 14, 15 and 16). Before LC-SRM analysis, unlabeled *DIGESTIF* peptides serving as quantification standards were added to the samples and the digests were desalted on Macrospin C18 columns (Harvard apparatus).

Biomarker quantification experiment

A plasma sample from a patient suffering from drug-induced liver injury was collected in the hepatology department at Grenoble university hospital. This patient was a member of a clinical research cohort established to assess new liver injury biomarkers. The research protocol was approved by the hospital's

institutional review board, and the patient provided written informed consent for participation in the study. The blood sample was taken upon hospital admission in a BD P100 tube (BD Biosciences, le Pont de Claix, France). It was centrifuged at 2200 g for 10 min to obtain plasma. This was immediately aliquoted and stored at -80 °C. Aliquots (3 µl each) of the patient's plasma sample were spiked with 0.5 µg (6 pmoles) of the *DIGESTIF* standard and were submitted to different proteolysis protocols, as described in Table 1 (Tests 1, 7, 8, 9, 10, 12 and 13). Plasma digests were either oxidized or alkylated as indicated for the different tests.

LC-SRM analyses

All SRM transition lists used to selectively monitor *DIGESTIF* peptides were generated using Skyline [22] (Table 2). LC-SRM runs were used to refine the SRM transition lists and to schedule acquisition. A similar procedure was followed when developing the SRM method to monitor the *DIGESTIF* peptides in combination with biomarker signature peptides in plasma. The following analyses were performed at the different sites:

Site 1

LC-SRM analyses were performed on a 4000QTrap hybrid triple quadrupole/ion trap mass spectrometer (AB Sciex) equipped with a TurboV source (AB Sciex) and controlled by Analyst software (version 1.5, AB Sciex). The instrument was equipped with an Ultima 3000 LC system (Thermo Scientific). Chromatography was performed using a two-solvent system

combining solvent A (2% acetonitrile, 0.1% formic acid) and solvent B (80% acetonitrile, 0.1% formic acid). 100 µg of protein digest (10 µl pick-up injection) was separated on a Kinetex XB-C18 column, 2.1 x 10 cm, 1.7 µm, 100 Å (Phenomenex, Le Pecq, France). Peptide separation was achieved using a linear gradient from 4% to 45% B over 25 min, and from 45% to 90% B in 5 min at a flow rate of 50 µl/min. MS data were acquired in positive mode with an ion spray voltage of 5500 V; curtain gas was used at 15 p.s.i. and the interface heater temperature was set to 375 °C. Collision exit, declustering and entrance potentials were set to 19, 12 and 55 V, respectively. The appropriate collision energy was calculated based on the following equations: CE (Volts) = 0.44*m/z + 4 for doubly-charged precursors and CE (Volts) = 0.5*m/z + 5 for triply-charged precursors. Analyses combined in the same run: (1) a precursor ion scan between 400-1400 m/z as a survey scan for Information Dependent Acquisition (IDA), (2) an Enhanced Product Ion (EPI) scan with a scan speed of 4000 amu/sec and a dynamic fill time for optimal MS/MS analysis, (3) an SRM acquisition with Q1 and Q3 quadrupoles operating at unit resolution. For scheduled SRM analyses, the acquisition time windows and target scan time were set to 90 sec and 1 sec, respectively. This corresponds, for chromatographic peaks with a base width of 30 sec, to acquisition of 30 points per LC peak.

Sites 2 and 3

At sites 2 and 3, SRM was performed on a TSQ Quantum Ultra EMR triple quadrupole mass spectrometer (Thermo Scientific) operated with Xcalibur 2.0.7

(Thermo Scientific). For each analysis, 1 µg of protein digest (1 µl pick-up injection) was injected via the associated nano-LC system (Eksigent). Samples were automatically injected into a 10-µl sample loop and loaded onto an analytical column. The analytical column (8 cm length × 75 µm internal diameter) was packed in-house with Magic C18 AQ beads (5 µm, 100 Å; Michrom Bioresources). Peptide mixtures were delivered to the analytical column at a flow rate of 500 nl/min of buffer A (5% acetonitrile, 0.2% formic acid) for 18 min and then eluted using a gradient of acetonitrile (10%–35%; 0.36%/min) in 0.2% formic acid at a flow rate of 250 nl/min. Collision energies (CE) were calculated according to the following formulas: CE = 0.034×m/z+3.314 (2+) and CE = 0.044×m/z+3.314 (3+). MS data were acquired in positive mode with an ion spray voltage of 1600 V and a capillary temperature of 270 °C. Scan speed was set to 20 ms per scan event, which resulted in a cycle time of 1.5 sec. For measurements, the Q1 resolution was set to 0.4 and the Q3 resolution to 0.7 at full width half maximum.

Data analysis

LC-SRM data were analyzed using Skyline [22]. Signals with a signal to noise ratio < 3 (*i.e.* peak height to background median < 3) or with obvious matrix interferences (experiments in biological matrices) were systematically excluded from the analysis. For absolute quantification of *DIGESTIF* peptides, coelution profiles of the labeled and unlabeled versions of each peptide were verified. Signals with a signal to noise ratio > 10 were considered for quantification.

Labeled/unlabeled peak area ratios were calculated for each SRM transition. Ratios obtained from the different SRM transitions were used to calculate the corresponding average peptide ratio. Peptide quantities in the processed sample were deduced from the ratios obtained for the different peptides and the quantities of spiked peptide standards.

RESULTS

Design of *DIGESTIF* standard

The major goal of the *DIGESTIF* project was to generate and test a universally applicable reference protein to assess the efficiency of trypsin digestion and LC-MS performance in bottom-up proteomics protocols. To achieve these goals, the *DIGESTIF* standard was designed as an isotope-labeled full-length protein which could be spiked into biological samples and digested with trypsin alongside endogenous proteins. The *DIGESTIF* standard included in its sequence the 11 peptides of the iRT peptide set previously described by Escher and colleagues. These peptides span a wide range of hydrophobicity and eluted at regular intervals using C18 reverse-phase columns and standard LC gradients [20].

To simulate the differing digestion properties encountered in a complex protein mixture, iRT peptide cleavage sites were flanked with amino acids to either favor or hamper trypsin digestion (Figure 1). These amino acids were chosen based on an in-depth study of trypsin cleavage rules [34]. Depending on di-amino acid motifs present at cleavage sites, the iRT peptides contained in the *DIGESTIF* standard were classified in 3 categories: those that would be easily released, those that would be released at a moderate rate and those that would be difficult to release (Supplementary Figure S2). The sequences corresponding to these flanked iRT peptides were inserted into the coding sequence of the soluble PBP2x protein scaffold from *Streptococcus pneumoniae* [44] (Figure 1),

and molecular modeling was used to determine the most suitable sequential arrangement of the iRT peptides in the PBP2x permissive regions (with respect to solubility and stability) based on three prototype designs (see experimental section). As proteolysis rate is also influenced by secondary and tertiary structure, a peptide from the β7 sheet of the PBP2x core (peptide NLVQPIVVGTGTK) was also included in the panel of indicator peptides.

The *DIGESTIF* construct was expressed in *E. coli* as a ([$^{13}C_6$, $^{15}N_4$] arginine and [$^{13}C_6$, $^{15}N_2$] lysine) isotopically labeled protein. Isotope incorporation was measured and found to be greater than 99% (Supplementary Figure S1). The purified, refolded and labeled protein standard will hereafter be referred to as *DIGESTIF* and is the form used in all the experiments described in this paper. The isotope labeling of *DIGESTIF* makes it possible to distinguish between iRT peptides released by proteolytic digestion of *DIGESTIF* and the commercially available, unlabeled iRT peptides [20].

Features and kinetics of *DIGESTIF* proteolysis

Before using the *DIGESTIF* standard in complex protein matrices, we characterized its behavior as sole sample constituent by measuring the peptides generated upon trypsin digestion. Pure *DIGESTIF* standard was subjected to digestion with trypsin at 1/100 (w/w) enzyme-to-protein ratio for 24 h at 37 °C. An oxidation step (transforming cysteine into cysteic acid) was included prior to LC-MS analysis for optimal detection of the missed-cleaved, cysteine-containing peptide TACK**DGLDAASYYAPVR** [45]. We chose to use cysteine oxidation rather than regular reduction/alkylation. As oxidation is performed after digestion, it

cannot have an impact on proteolysis. After peptide oxidation, the digest was analyzed by LC-MS/MS on an Orbitrap mass spectrometer (Thermo Scientific). The peptides identified covered 82% of the *DIGESTIF* sequence (Figure 2A), including all iRT peptides. Among the observed iRT peptides, the ADVTPADFSEWSK peptide (iRT peptide h) was detected with an oxidized tryptophan residue. This is probably due to the oxidation step. As tryptophan oxidation can yield several oxidation products, this peptide was not monitored in subsequent experiments. A few iRT peptides were detected, at low intensities, in missed-cleaved forms. These consisted of the following: CAR**YILAGVENSK** (iRT peptide d), **YILAGVENSK**TAAR (iRT peptide d) (data not shown), and the expected - and by far the most prevalent - TACK**DGLDAASYYAPVR** (iRT peptide g) (Figure 2B).

Next, to determine the kinetics of iRT peptide release during *DIGESTIF* proteolysis, pure *DIGESTIF* standard was submitted to tryptic digestion under differing conditions. Two different enzyme-to-protein ratios were tested: (i) 1/100 (w/w) (standard concentration), and (ii) 1/1000 (w/w) which is 10-times below supplier's recommendations. We used this low concentration to simulate sub-optimal sample processing conditions. In addition to the variation in enzyme concentrations, the digestion reaction was stopped at different time points to determine the kinetics of peptide release. Analysis of the peptides released was performed using LC-SRM to enhance detection sensitivity (Figure 3). This is crucial, as at early stages of the digestion reaction peptides will only be present in minute amounts. In these assays, all iRT peptides were quantified using a

label-free strategy based on the signal provided by the best transition (Supplementary Table ST1). Characterization of *DIGESTIF* proteolysis was further refined and improved by quantifying four selected peptides based on signal comparison with unlabeled peptide standards. The four selected peptides were: iRT peptide i (GTFIIDPGGVIR, predicted to be easily cleaved), iRT peptide g (DGLDAASYYAPVR, predicted to be difficult to digest), missed-cleaved iRT peptide g (TACK**DGLDAASYYAPVR**) and the peptide from the PBP2x protein core (peptide NLVQPIVVGTGTK) (Figure 3B).

With a trypsin-to-protein ratio of 1/100, all iRT peptides flanked with "digestion-promoting" cleavage sites (Figure 1) were rapidly released. Steady state levels of these peptides were attained after 5 hours of incubation. iRT peptides predicted to be difficult to cleave, *i.e.* iRT peptides c (VEATFGVDESNAK) and g (DGLDAASYYAPVR) did not reach a plateau level, even after 24 h of proteolysis (Figure 3B and Supplementary Table ST1). The signal intensity for the missed-cleaved version of peptide g, TACK**DGLDAASYYAPVR**, peaked at 5 h of digestion and decreased thereafter (when the correctly cleaved peptide g started to become detectable); while the PBP2x core peptide (NLVQPIVVGTGTK) displayed regular, slow digestion kinetics in these conditions, probably due to its protected position in the PBP2x protein core [44]. In the alternative enzymatic conditions, when using a trypsin-to-protein ratio of 1/1000, even with prolonged digestion times digestion did not reach completion for most of the peptides monitored (Figure 3, Supplementary Table ST1). Based on these results, we conclude that the release of iRT peptides

(which show distinct digestion kinetics), directly reflects protein susceptibility to trypsin proteolysis, as a function of primary sequence. The release of the PBP2x core peptide evidences the influence of protein structure on trypsin activity. In summary, the *DIGESTIF* standard can be used as a model reflecting the digestion properties of a broad range of proteins.

Characterizing *DIGESTIF* proteolysis in human serum

Our initial aim in designing the *DIGESTIF* standard was to improve the reliability of MS-based analyses, in particular for biomarker evaluation studies. It was thus necessary to assess how *DIGESTIF* performed in the complex biofluids commonly used in this type of study, such as serum or plasma. For this purpose, serum aliquots (6 µL each) from a healthy donor were spiked with 1 µg of *DIGESTIF* standard. Considering molar abundance, *DIGESTIF* standard ranked below the 30 to 50 most abundant plasma proteins. Samples were submitted to proteolysis using eight reaction conditions (Figure 4). These included both classical procedures and sub-optimal proteolysis conditions that would reflect possible mistakes made when executing the protocol. The resulting peptide samples were analyzed by LC-SRM. All the peptides released were quantified using a label-free strategy (Supplementary Table ST2). Three *DIGESTIF* peptides (iRT peptides g and i, and PBP2x core peptide) were also quantified using unlabeled peptide standards to precisely evaluate the efficiency of proteolysis across the different conditions (Figure 4). These experiments showed that the release of iRT peptides from *DIGESTIF* was reduced by a low trypsin-to-

protein ratio, sub-optimal incubation temperature (30 °C or 23 °C) or acidic pH. In line with results reported in previous studies [24,28], we found that the best proteolysis protocol was based on the use of a trypsin/Lys-C enzymatic mixture and urea for protein denaturation (Figure 4, Test 9). In these conditions, all iRT peptides and the PBP2x core peptide were more readily detected than when samples were digested with trypsin alone (Test 1) (Figure 4 and Supplementary Table ST2). In quantitative terms, this improvement corresponds to an increase in peptide abundances by between 1.2-fold (PBP2x core peptide) and 22.4-fold (iRT peptide g) compared to Test 1 (Figure 4). Interestingly, treating the serum sample with TCEP and iodoacetamide before trypsin digestion (Test 10) also promoted the release, and thus detection, of iRT peptides from *DIGESTIF* (Figure 4 and Supplementary Table ST2), although the improvement was not as marked as in the trypsin/Lys-C mix plus urea conditions (Test 9). From these results, we conclude that *DIGESTIF* can be used to evaluate digestion efficiency in the complex matrices typically used in biomarker evaluation studies, and that it is useful in identifying conditions where proteolysis is incomplete. It also allows to select the optimal digestion protocol.

Characterizing *DIGESTIF* proteolysis in human urine and mouse liver lysate

As the peptides generated from *DIGESTIF* are artificial, this standard can potentially be used in combination with any biological sample. To demonstrate this applicability, we went on to characterize its performance in two additional biological matrices: human urine and mouse liver lysate. In these tests, both

samples were spiked with *DIGESTIF* standard before being submitted to different proteolysis protocols (Figure 5). The iRT peptides g and i, and PBP2x core peptide were quantified using LC-SRM and unlabeled peptide standards. In human urine, these iRT peptides were more efficiently released when trypsin was used in combination with reduction/alkylation (Test 10, Figure 5, Supplementary Table ST2). As it was impossible to apply the proteolysis protocol recommended by the trypsin/Lys-C mix manufacturer (Test 9) with urine samples as it would have led to excessive sample dilution, we used the trypsin/Lys-C mix in combination with FASP (Table 1). This latter protocol was also found efficient to release the three targeted iRT peptides (Figure 5). In summary, our results confirm that a protocol combining reduction/alkylation and trypsin digestion is appropriate for sensitive detection of urinary proteins [46].

Mouse liver lysate is a more complex and more concentrated matrix than urine. In this sample, only 8 out of the 10 monitored iRT peptides could be detected (data not shown). The released amounts of indicator peptides (iRT peptides g, i and the PBP2x core peptide) indicate that the proteolysis using trypsin/Lys-C mix clearly outperformed the other protocols tested (Test 9, Figure 5, Supplementary Table ST2).

In conclusion, these experiments confirm the importance of adapting and assessing proteolysis protocols for each type of biological matrix [15,16,24,28,30]. The *DIGESTIF* standard undoubtedly facilitates protocol comparison during these tests.

Improving MS-based biomarker studies using *DIGESTIF*

To test the suitability of *DIGESTIF* in the context of clinical biomarker detection, three biomarkers were investigated in a plasma sample from a patient with drug-induced liver injury. The biomarkers investigated were the following: alanine aminotransferase 1 (ALAT1) - which is routinely investigated in clinical settings -, liver fatty acid binding protein (L-FABP), and alcohol dehydrogenase 1 (ADH1) - which have both been described as potential liver injury biomarkers [47,48]. The patient's plasma sample was spiked with the *DIGESTIF* standard before sample digestion to assess proteolysis efficiency. As in previous experiments, seven commonly used digestion protocols were compared (Table 1, Tests 1, 7, 8, 9, 10, 12 and 13), and the resulting samples were analyzed using LC-SRM (Figure 6). Label-free data analysis revealed two main classes of *DIGESTIF* peptides: (i) peptides that were readily detected across all proteolysis conditions (iRT peptides b, c, d, i and the PBP2x core peptide); and (ii) peptides that were preferentially observed in the conditions applied in Test 9 and Test 13 (iRT peptides a, e, f, g, k and l) (Figure 7, Supplementary Table ST3). Besides monitoring iRT peptide release, we also determined which protocol was most efficient for the simultaneous and sensitive detection of the 3 targeted biomarkers. This turned out to be Test 9 (trypsin/Lys C with urea; Figure 6 lower panel), confirming our assumption that some biomarkers (here ADH1, ALAT1) are very sensitive to variations in proteolysis conditions (Figure 7) [17,38,49]. Together, these results demonstrate that the *DIGESTIF* standard, particularly iRT

peptides a, e, f, g, k and l, can be used to identify optimal proteolysis protocols for sensitive biomarker detection.

Multi-site assessment of *DIGESTIF* in a serum matrix

MS-based biomarker assays are destined to be used in many laboratories and at different times, conditions which are very difficult to standardize. Therefore, we next examined how the *DIGESTIF* standard could be used to assess digestion efficiency and reproducibility across different laboratories. For these experiments, *DIGESTIF* (1 µg) was spiked into a serum sample (6 µl) before proteolysis in three different laboratories. The same protocols (Tests 1, 7, 8, 11 and 12, Table 1) and the same reagents (but different batches) were used in all three laboratories, and samples were analyzed by LC-SRM. Quantification data obtained for the peptides released were then cross-compared (Figure 8, Supplementary Table ST4). Once again, proteolysis efficiency was seen to vary extensively between protocols, but interestingly it also varied quite a lot between laboratories, particularly in sub-optimal conditions. However, as previously noticed, the “double-digestion” protocol, using a trypsin/Lys-C mix in the presence of urea (Test 11), was clearly identified as the best for human serum (Figure 8, Supplementary Table ST4). In these conditions, compared to standard trypsin proteolysis (Test 1), across the 3 laboratories, iRT peptides i (GTFIIDPGGVIR) and g (DGLDAASYYAPVR) were released 5.2 (± 5.7) and 6.0 (± 1.4) times more effectively, respectively. In conclusion, *DIGESTIF* clearly

identified the most efficient protocol when comparing inter-laboratory results. It also unequivocally highlighted inter-laboratory variations in digestion efficiency.

DISCUSSION

In this article we present *DIGESTIF,* a generic standard for use in bottom-up proteomics analyses. This standard can be used to verify the quality of all the steps in the sample treatment workflow from digestion to LC-MS analysis as part of intra- and inter-laboratory quality control.

The *DIGESTIF* standard consists in a PBP2x protein scaffold [40], to which the sequences corresponding to the set of iRT peptides [20] have been added. This scaffold was chosen because previous experiments had demonstrated that several of its regions were structurally permissive for engineering. Thus, removal, substitution or insertion of amino acids in these regions has little impact on the core structure, solubility and functionality of the engineered PBP2x protein variants [40]. The PBP2x scaffold thus serves several purposes in the context of the *DIGESTIF* standard: (1) it carries and exposes iRT peptides [20] to trypsin digestion; (2) it favors solubility and stability of the whole standard; (3) it facilitates the use of the standard by providing a single reagent.

Bottom-up proteomics workflows involve three main steps: proteolysis, liquid chromatography, and mass spectrometry analysis. Proteolysis is an essential step. However, conversion of proteins into peptides remains difficult to standardize and is often not reproducible. Although Piehowski and coworkers [50] estimate that protein digestion only accounts for 3.1% of intra-laboratory technical variability, several inter-laboratory studies (including this work) have indicated a higher variability of trypsin digestion when the same sample is treated at different sites [17,51]. This is important to take into consideration as variations in

proteolysis will have a direct impact on peptide recovery, and thus detection sensitivity. Quantitative experiments using labeled proteins as quantification references (SILAC, PSAQ) [38,52] will not be dramatically affected by these variations, but in label-free or AQUA experiments [24], digestion yields will directly affect quantification accuracy, and thus hamper attempts to discover or use protein biomarkers. Percy and colleagues [23] were the first to present a dedicated tool to standardize and attempt to benchmark the entire proteomics-based biomarker detection workflow, including proteolysis. However, this tool was only adapted for use in human plasma samples. The *DIGESTIF* standard presented here overcomes this limitation as it can assess the general efficiency of proteolysis, whatever the composition and properties of the sample. *DIGESTIF* combines several quality control aspects in a single reagent whose main purpose is to check the quality and efficiency of digestion. This was made possible through careful design. Indeed, unlike most other standards (e.g. QconCATs), *DIGESTIF* is not optimized for efficient digestion. Rather, it releases indicator peptides progressively, directly reflecting the overall digestion performance in complex and heterogenous protein samples. Hence, *DIGESTIF* mimics the properties of a broad range of proteins and, as presented here, can be used to establish and optimize the digestion protocol for any sample, regardless of the properties of the matrix or specific target proteins, or the type of MS-based detection employed. Once an optimized digestion protocol has been established, *DIGESTIF* can be used to monitor digestion efficiency over multiple experiments.

To simultaneously enable quality control of the LC step with our standard, we chose to integrate iRT peptides into its sequence. These peptides were specifically designed to distribute evenly over a standard reverse-phase LC gradient [20]. Alternatives to the iRT peptide set could have been used. Indeed, we also tested the QCAL peptides presented by Eyers and coworkers [19] as part of our initial selection. However, as five of the QCAL peptides co-eluted in our gradient conditions (data not shown), these peptides were not deemed suitable for our purposes. Another advantage of the iRT peptide set is that, with SRM measurements, automated target validation is possible using mProphet or Skyline [21,22].

In addition to monitoring digestion efficiency and reproducibility, the *DIGESTIF* standard will be particularly useful in the context of biomarker studies, where false negative results can be problematic. By indicating the degree of digestion, *DIGESTIF* can be used to rate the quality of sample processing based on reliable and objective criteria (e.g. release of iRT peptide l, see Figure 7). This will help to exclude poorly-digested samples from evaluation. Likewise, *DIGESTIF* will make data comparison between different studies more reliable, while also helping to eliminate false positives that might simply be due to differences between digestion protocols. This is also essential for comparative (especially label-free) proteomics approaches where conclusions will be drawn based on abundance differences between two or more experimental states.

The *DIGESTIF* standard presented in this publication will be made available to researchers and is expected to be commercially available as a

validated product for the broad proteomics community. More extensive experience with the *DIGESTIF* standard will help to refine and validate its use in a range of different contexts. We believe that this standard will help researchers to select robust proteolysis protocols, and to define quality control criteria for biomarker experiments based on defined relationships (e.g. ratios) of a subset of *DIGESTIF* peptides. This will be further supported by improved versions of the standard which will include new features to check for protein modifications (glycosylations), for chemical reactions occurring during sample processing, or use other proteases than trypsin.

In conclusion, *DIGESTIF* can be used to: (i) benchmark digestion reagents (*i.e.* proteases) before establishing digestion protocols [25,31], (ii) select the best conditions for sample proteolysis [26,28,29], (iii) check the efficiency of digestion, (iv) identify poorly digested samples and correct for variations in digestion efficiency, if necessary. We believe that *DIGESTIF* has great potential to enhance the quality and reproducibility of bottom-up proteomics experiments.

REFERENCES

(1) Bensimon, A.; Heck, A. J.; Aebersold, R.: Mass spectrometry-based proteomics and network biology. *Annu Rev Biochem* **2012**, *81*, 379-405.

(2) Gillette, M. A.; Carr, S. A.: Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry. *Nature methods* **2013**, *10*, 28-34.

(3) Taguchi, A.; Hanash, S. M.: Unleashing the power of proteomics to develop blood-based cancer markers. *Clinical chemistry* **2013**, *59*, 119-26.

(4) Beck, A.; Sanglier-Cianferani, S.; Van Dorsselaer, A.: Biosimilar, biobetter, and next generation antibody characterization by mass spectrometry. *Analytical chemistry* **2012**, *84*, 4637-46.

(5) Ahrens, C. H.; Brunner, E.; Qeli, E.; Basler, K.; Aebersold, R.: Generating and navigating proteome maps using mass spectrometry. *Nat Rev Mol Cell Biol* **2010**, *11*, 789-801.

(6) Ong, S. E.; Mann, M.: Mass spectrometry-based proteomics turns quantitative. *Nat Chem Biol* **2005**, *1*, 252-62.

(7) Tipton, J. D.; Tran, J. C.; Catherman, A. D.; Ahlf, D. R.; Durbin, K. R.; Kelleher, N. L.: Analysis of intact protein isoforms by mass spectrometry. *J Biol Chem* **2011**, *286*, 25451-8.

(8) Stoevesandt, O.; Taussig, M. J.: Affinity proteomics: the role of specific binding reagents in human proteome analysis. *Expert Rev Proteomics* **2012**, *9*, 401-14.

(9) Tabb, D. L.; Vega-Montoto, L.; Rudnick, P. A.; Variyath, A. M.; Ham, A. J.; Bunk, D. M.; Kilpatrick, L. E.; Billheimer, D. D.; Blackman, R. K.; Cardasis, H. L.; Carr, S. A.; Clauser, K. R.; Jaffe, J. D.; Kowalski, K. A.; Neubert, T. A.; Regnier, F. E.; Schilling, B.; Tegeler, T. J.; Wang, M.; Wang, P.; Whiteaker, J. R.; Zimmerman, L. J.; Fisher, S. J.; Gibson, B. W.; Kinsinger, C. R.; Mesri, M.; Rodriguez, H.; Stein, S. E.; Tempst, P.; Paulovich, A. G.; Liebler, D. C.; Spiegelman, C.: Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *Journal of proteome research* **2010**, *9*, 761-76.

(10) Carr, S. A.; Abbatiello, S. E.; Ackermann, B. L.; Borchers, C.; Domon, B.; Deutsch, E. W.; Grant, R. P.; Hoofnagle, A. N.; Huttenhain, R.; Koomen, J. M.; Liebler, D. C.; Liu, T.; MacLean, B.; Mani, D. R.; Mansfield, E.; Neubert, H.; Paulovich, A. G.; Reiter, L.; Vitek, O.; Aebersold, R.; Anderson, L.; Bethem, R.; Blonder, J.; Boja, E.; Botelho, J.; Boyne, M.; Bradshaw, R. A.; Burlingame, A. L.; Chan, D.; Keshishian, H.; Kuhn, E.; Kinsinger, C.; Lee, J. S.; Lee, S. W.; Moritz, R.; Oses-Prieto, J.; Rifai, N.; Ritchie, J.; Rodriguez, H.; Srinivas, P. R.; Townsend, R. R.; Van Eyk, J.; Whiteley, G.; Wiita, A.; Weintraub, S.: Targeted peptide measurements in biology and medicine: best practices for mass spectrometry-based assay development using a fit-for-purpose approach. *Molecular & cellular proteomics : MCP* **2014**, *13*, 907-17.

(11) Domon, B.; Aebersold, R.: Options and considerations when selecting a quantitative proteomics strategy. *Nat Biotechnol* **2010**, *28*, 710-21.

(12) Ivanov, A. R.; Colangelo, C. M.; Dufresne, C. P.; Friedman, D. B.; Lilley, K. S.; Mechtler, K.; Phinney, B. S.; Rose, K. L.; Rudnick, P. A.; Searle, B. C.; Shaffer, S. A.; Weintraub, S. T.: Interlaboratory studies and initiatives developing standards for proteomics. *Proteomics* **2013**, *13*, 904-9.

(13) Shuford, C. M.; Sederoff, R. R.; Chiang, V. L.; Muddiman, D. C.: Peptide production and decay rates affect the quantitative accuracy of protein cleavage isotope dilution mass spectrometry (PC-IDMS). *Molecular & cellular proteomics : MCP* **2012**, *11*, 814-23.

(14) Burkhart, J. M.; Premsler, T.; Sickmann, A.: Quality control of nano-LC-MS systems using stable isotope-coded peptides. *Proteomics* **2011**, *11*, 1049-57.

(15) Proc, J. L.; Kuzyk, M. A.; Hardie, D. B.; Yang, J.; Smith, D. S.; Jackson, A. M.; Parker, C. E.; Borchers, C. H.: A quantitative study of the effects of chaotropic agents, surfactants, and solvents on the digestion efficiency of human plasma proteins by trypsin. *Journal of proteome research* **2010**, *9*, 5422-37.

(16) Burkhart, J. M.; Schumbrutzki, C.; Wortelkamp, S.; Sickmann, A.; Zahedi, R. P.: Systematic and quantitative comparison of digest efficiency and specificity reveals the impact of trypsin quality on MS-based proteomics. *Journal of proteomics* **2012**, *75*, 1454-62.

(17) Addona, T. A.; Abbatiello, S. E.; Schilling, B.; Skates, S. J.; Mani, D. R.; Bunk, D. M.; Spiegelman, C. H.; Zimmerman, L. J.; Ham, A. J.; Keshishian, H.; Hall, S. C.; Allen, S.; Blackman, R. K.; Borchers, C. H.; Buck, C.; Cardasis, H.

L.; Cusack, M. P.; Dodder, N. G.; Gibson, B. W.; Held, J. M.; Hiltke, T.; Jackson, A.; Johansen, E. B.; Kinsinger, C. R.; Li, J.; Mesri, M.; Neubert, T. A.; Niles, R. K.; Pulsipher, T. C.; Ransohoff, D.; Rodriguez, H.; Rudnick, P. A.; Smith, D.; Tabb, D. L.; Tegeler, T. J.; Variyath, A. M.; Vega-Montoto, L. J.; Wahlander, A.; Waldemarson, S.; Wang, M.; Whiteaker, J. R.; Zhao, L.; Anderson, N. L.; Fisher, S. J.; Liebler, D. C.; Paulovich, A. G.; Regnier, F. E.; Tempst, P.; Carr, S. A.: Multi-site assessment of the precision and reproducibility of multiple reaction monitoring-based measurements of proteins in plasma. *Nat Biotechnol* **2009**, *27*, 633-41.

(18) Paulovich, A. G.; Billheimer, D.; Ham, A. J.; Vega-Montoto, L.; Rudnick, P. A.; Tabb, D. L.; Wang, P.; Blackman, R. K.; Bunk, D. M.; Cardasis, H. L.; Clauser, K. R.; Kinsinger, C. R.; Schilling, B.; Tegeler, T. J.; Variyath, A. M.; Wang, M.; Whiteaker, J. R.; Zimmerman, L. J.; Fenyo, D.; Carr, S. A.; Fisher, S. J.; Gibson, B. W.; Mesri, M.; Neubert, T. A.; Regnier, F. E.; Rodriguez, H.; Spiegelman, C.; Stein, S. E.; Tempst, P.; Liebler, D. C.: Interlaboratory study characterizing a yeast performance standard for benchmarking LC-MS platform performance. *Molecular & cellular proteomics : MCP* **2010**, *9*, 242-54.

(19) Eyers, C. E.; Simpson, D. M.; Wong, S. C.; Beynon, R. J.; Gaskell, S. J.: QCAL--a novel standard for assessing instrument conditions for proteome analysis. *J Am Soc Mass Spectrom* **2008**, *19*, 1275-80.

(20) Escher, C.; Reiter, L.; MacLean, B.; Ossola, R.; Herzog, F.; Chilton, J.; MacCoss, M. J.; Rinner, O.: Using iRT, a normalized retention time for more targeted measurement of peptides. *Proteomics* **2012**, *12*, 1111-21.

(21) Reiter, L.; Rinner, O.; Picotti, P.; Huttenhain, R.; Beck, M.; Brusniak, M. Y.; Hengartner, M. O.; Aebersold, R.: mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nature methods* **2011**, *8*, 430-5.

(22) MacLean, B.; Tomazela, D. M.; Shulman, N.; Chambers, M.; Finney, G. L.; Frewen, B.; Kern, R.; Tabb, D. L.; Liebler, D. C.; MacCoss, M. J.: Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **2010**, *26*, 966-8.

(23) Percy, A. J.; Chambers, A. G.; Yang, J.; Jackson, A. M.; Domanski, D.; Burkhart, J.; Sickmann, A.; Borchers, C. H.: Method and platform standardization in MRM-based quantitative plasma proteomics. *Journal of proteomics* **2013**, *95*, 66-76.

(24) Brownridge, P.; Beynon, R. J.: The importance of the digest: proteolysis and absolute quantification in proteomics. *Methods* **2011**, *54*, 351-60.

(25) Bunkenborg, J.; Espadas, G.; Molina, H.: Cutting edge proteomics: benchmarking of six commercial trypsins. *Journal of proteome research* **2013**, *12*, 3631-41.

(26) Chiva, C.; Ortega, M.; Sabido, E.: Influence of the Digestion Technique, Protease, and Missed Cleavage Peptides in Protein Quantitation. *Journal of proteome research* **2014**, *13*, 3979-86.

(27) Fonslow, B. R.; Stein, B. D.; Webb, K. J.; Xu, T.; Choi, J.; Park, S. K.; Yates, J. R., 3rd: Digestion and depletion of abundant proteins improves proteomic coverage. *Nature methods* **2013**, *10*, 54-6.

(28) Glatter, T.; Ludwig, C.; Ahrne, E.; Aebersold, R.; Heck, A. J.; Schmidt, A.: Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. *Journal of proteome research* **2012**, *11*, 5145-56.

(29) Hildonen, S.; Halvorsen, T. G.; Reubsaet, L.: Why less is more when generating tryptic peptides in bottom-up proteomics. *Proteomics* **2014**, *14*, 2031-41.

(30) Switzar, L.; Giera, M.; Niessen, W. M. A.: Protein digestion: an overview of the available techniques and recent developments. *Journal of proteome research* **2013**, *12*, 1067-77.

(31) Walmsley, S. J.; Rudnick, P. A.; Liang, Y.; Dong, Q.; Stein, S. E.; Nesvizhskii, A. I.: Comprehensive analysis of protein digestion using six trypsins reveals the origin of trypsin as a significant source of variability in proteomics. *Journal of proteome research* **2013**, *12*, 5666-80.

(32) Keil: Specificity of Proteolysis. Springer Verlag. Berlin. **1992**.

(33) Olsen, J. V.; Ong, S. E.; Mann, M.: Trypsin cleaves exclusively C-terminal to arginine and lysine residues. *Molecular & cellular proteomics : MCP* **2004**, *3*, 608-14.

(34) Rodriguez, J.; Gupta, N.; Smith, R. D.; Pevzner, P. A.: Does trypsin cut before proline? *Journal of proteome research* **2008**, *7*, 300-5.

(35) Fannes, T.; Vandermarliere, E.; Schietgat, L.; Degroeve, S.; Martens, L.; Ramon, J.: Predicting tryptic cleavage from proteomics data using decision tree ensembles. *Journal of proteome research* **2013**, *12*, 2253-9.

(36) Cutillas, P. R.; Geering, B.; Waterfield, M. D.; Vanhaesebroeck, B.: Quantification of gel-separated proteins and their phosphorylation sites by LC-MS using unlabeled internal standards: analysis of phosphoprotein dynamics in a B cell lymphoma cell line. *Molecular & cellular proteomics : MCP* **2005**, *4*, 1038-51.

(37) Beynon, R. J.; Doherty, M. K.; Pratt, J. M.; Gaskell, S. J.: Multiplexed absolute quantification in proteomics using artificial QCAT proteins of concatenated signature peptides. *Nature methods* **2005**, *2*, 587-9.

(38) Brun, V.; Dupuis, A.; Adrait, A.; Marcellin, M.; Thomas, D.; Court, M.; Vandenesch, F.; Garin, J.: Isotope-labeled Protein Standards: Toward Absolute Quantitative Proteomics. *Molecular & cellular proteomics : MCP* **2007**, *6*, 2139-2149.

(39) Rivers, J.; Simpson, D. M.; Robertson, D. H.; Gaskell, S. J.; Beynon, R. J.: Absolute multiplexed quantitative analysis of protein expression during muscle development using QconCAT. *Molecular & cellular proteomics : MCP* **2007**, *6*, 1416-27.

(40) Dideberg, O. V., T. Mouz, N.: Patent EP1521771 Streptococcus pneumoniae PBP2x mini-protein and uses thereof.

(41) Matic, I.; Jaffray, E. G.; Oxenham, S. K.; Groves, M. J.; Barratt, C. L.; Tauro, S.; Stanley-Wall, N. R.; Hay, R. T.: Absolute SILAC-compatible

expression strain allows Sumo-2 copy number determination in clinical samples. *Journal of proteome research* **2011**, *10*, 4869-75.

(42) Manza, L. L.; Stamer, S. L.; Ham, A. J.; Codreanu, S. G.; Liebler, D. C.: Sample preparation and digestion for proteomic analyses using spin filters. *Proteomics* **2005**, *5*, 1742-5.

(43) Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M.: Universal sample preparation method for proteome analysis. *Nature methods* **2009**, *6*, 359-62.

(44) Pares, S.; Mouz, N.; Petillot, Y.; Hakenbeck, R.; Dideberg, O.: X-ray structure of Streptococcus pneumoniae PBP2x, a primary penicillin target enzyme. *Nat Struct Biol* **1996**, *3*, 284-9.

(45) Jaquinod, M.; Trauchessec, M.; Huillet, C.; Louwagie, M.; Lebert, D.; Picard, G.; Adrait, A.; Dupuis, A.; Garin, J.; Brun, V.; Bruley, C.: Mass spectrometry-based absolute protein quantification: PSAQ strategy makes use of "noncanonical" proteotypic peptides. *Proteomics* **2012**, *12*, 1217-21.

(46) Selevsek, N.; Matondo, M.; Sanchez Carbayo, M.; Aebersold, R.; Domon, B.: Systematic quantification of peptides/proteins in urine using selected reaction monitoring. *Proteomics* **2011**, *11*, 1135-47.

(47) Pelsers, M. M.; Morovat, A.; Alexander, G. J.; Hermens, W. T.; Trull, A. K.; Glatz, J. F.: Liver fatty acid-binding protein as a sensitive serum marker of acute hepatocellular damage in liver transplant recipients. *Clinical chemistry* **2002**, *48*, 2055-7.

(48) Collins, B. C.; Miller, C. A.; Sposny, A.; Hewitt, P.; Wells, M.; Gallagher, W. M.; Pennington, S. R.: Development of a pharmaceutical hepatotoxicity biomarker panel using a discovery to targeted proteomics approach. *Molecular & cellular proteomics : MCP* **2012**, *11*, 394-410.

(49) Whiteaker, J. R.; Zhao, L.; Anderson, L.; Paulovich, A. G.: An automated and multiplexed method for high throughput peptide immunoaffinity enrichment and multiple reaction monitoring mass spectrometry-based quantification of protein biomarkers. *Molecular & cellular proteomics : MCP* **2010**, *9*, 184-96.

(50) Piehowski, P. D.; Petyuk, V. A.; Orton, D. J.; Xie, F.; Moore, R. J.; Ramirez-Restrepo, M.; Engel, A.; Lieberman, A. P.; Albin, R. L.; Camp, D. G.; Smith, R. D.; Myers, A. J.: Sources of technical variability in quantitative LC-MS proteomics: human brain tissue sample analysis. *Journal of proteome research* **2013**, *12*, 2128-37.

(51) Kuhn, E.; Whiteaker, J. R.; Mani, D. R.; Jackson, A. M.; Zhao, L.; Pope, M. E.; Smith, D.; Rivera, K. D.; Anderson, N. L.; Skates, S. J.; Pearson, T. W.; Paulovich, A. G.; Carr, S. A.: Interlaboratory evaluation of automated, multiplexed Peptide immunoaffinity enrichment coupled to multiple reaction monitoring mass spectrometry for quantifying proteins in plasma. *Molecular & cellular proteomics : MCP* **2012**, *11*, M111 013854.

(52) Ong, S. E.; Blagoev, B.; Kratchmarova, I.; Kristensen, D. B.; Steen, H.; Pandey, A.; Mann, M.: Stable isotope labeling by amino acids in cell culture,

SILAC, as a simple and accurate approach to expression proteomics. *Molecular & cellular proteomics : MCP* **2002**, *1*, 376-86.

ACKNOWLEDGEMENTS

We thank Audrey Lebas for molecular biology experiments, Pauline Maes for biomarker detection experiments, the team at EDyP Service for LC-MS/MS analyses, Dr. Michel Jaquinod for scientific support, and Maighread Gallagher-Gambarelli for editing services. We are also grateful to Dr. Leila El Kholy and Pr. Jean-Pierre Zarski for providing clinical samples. The Functional Genomics Centre, Zurich, is acknowledged for technical support. This project was supported by grants from the 7th Framework Programme of the European Union (Contract no. 262067- PRIME-XS) and Investissement d'Avenir Infrastructures Nationales en Biologie et Santé program (ProFI project, ANR-10-INBS-08). SG was supported by a grant from the Swiss National Science Foundation (Marie-Heim Voegtlin, PMPDP3_122836).

FIGURE LEGENDS

Figure 1. Design of the *DIGESTIF* standard

(**A**) The *DIGESTIF* protein was assembled using iRT peptides and a protein scaffold called PBP2x mini [40] from which N-terminal sequences (grey) have been removed. iRT peptide cleavage sites were flanked by amino acids that are known to influence digestion efficiency, as determined by Rodriguez and coworkers (supplementary Figure S2). Based on di-amino acid motifs surrounding cleavage sites, iRT peptides were designed to fall into one of 3 categories: sequences in green correspond to easily released peptides, orange lettering indicates moderately released peptides, and red peptides are classed as difficult to cleave. (**B**) In the final *DIGESTIF* protein the iRT peptides are distributed over three predetermined insertion sites in the PBP2x mini sequence. The optimal placement of the iRT peptides was determined by molecular modeling (see Experimental section for details). A hexahistidine tag (blue) inserted at the extreme N-terminus was used to purify the recombinant protein. (**C**) Full amino acid sequence of the *DIGESTIF* protein. PBP2x mini sequences are indicated in grey, iRT peptides in green, orange or red.

Figure 2. LC-MS/MS and LC-SRM analysis of *DIGESTIF* standard.

Pure *DIGESTIF* was submitted to in-solution trypsin digestion followed by oxidation. The digest was analyzed using LC-MS/MS on an Orbitrap mass spectrometer equipped with a nanoLC chromatography system. MS/MS data

were processed automatically using Mascot Daemon software (version 2.3, Matrix Science). Peptides were identified using a custom-built database containing the *DIGESTIF* sequence (**A**). The peptides identified from the *DIGESTIF* sequence are shown in blue. Boxes indicate the iRT peptides and the PBP2x core peptide. (**B**) The digest was also used to develop a scheduled LC-SRM method targeting iRT peptides, the missed-cleaved iRT peptide g and the PBP2x core peptide. The extracted ion chromatogram for the corresponding SRM transitions is presented.

Figure 3. Time-course analysis of *DIGESTIF* proteolysis.

Samples containing 2 µg of pure *DIGESTIF* standard were submitted to trypsin digestion either with a 1/100 (w/w) or a 1/1000 (w/w) enzyme-to-protein ratio for different periods of time. Completion of digestion was monitored by SDS-PAGE analysis (1 µg of hydrolyzed *DIGESTIF* loaded) and silver staining (**A**) or quantitative LC-SRM analysis (**B**). Absolute quantities of 4 *DIGESTIF* peptides (iRT peptides i and g, missed-cleaved iRT peptide g and PBP2x core peptide) were estimated using LC-SRM and unlabeled peptide standards. The quantities were plotted as a function of the duration of digestion. Each data point corresponds to the mean ± standard deviation (n = 3 SRM transitions per peptide, see also Supplementary Table ST1).

Figure 4. *DIGESTIF* proteolysis in serum.

Serum samples (6 µl aliquots) were spiked with *DIGESTIF* standard (1 µg) and submitted to 8 different digestion protocols. These included 3 classical procedures (Tests 1, 9 and 10) and 5 sub-optimal proteolysis conditions that would reflect mistakes made when executing the protocol (Tests 2 to 6). The efficiency of proteolysis was assessed by LC-SRM to determine the quantities of 3 *DIGESTIF* peptides (iRT peptides i and g, and PBP2x core peptide) after adding unlabeled peptides (AQUA) as quantification standards. Each bar corresponds to the mean ± standard deviation, calculated from 3 technical replicates (see also Supplementary Table ST2).

Figure 5. *DIGESTIF* proteolysis in human urine and mouse liver lysate.

The human urine and mouse liver lysate samples were spiked with *DIGESTIF* standard prior to digestion using different protocols (Tests 1, 9, 10, 14, 15 or 16). Samples were then analyzed by LC-SRM targeting iRT peptides i and g, and the PBP2x core peptide. The amount of released peptide was measured using unlabeled peptide standard and is reported in the different digestion conditions. To increase detection sensitivity, with urine samples, LC-SRM analysis was performed using a QTrap6500 mass spectrometer. Due to the reduced m/z range (400-1000) provided by this machine, some transitions had to be changed (supplementary Table ST2).

Figure 6. LC-SRM chromatograms of *DIGESTIF* standard peptides and liver injury biomarker peptides in a patient plasma sample.

A plasma sample from a patient suffering from drug-induced liver injury was spiked with *DIGESTIF* standard. Aliquots of the plasma sample were digested by applying seven different proteolysis protocols (Tests 1, 7, 8, 9, 10, 12 and 13, see Table 1) and analyzed by LC-SRM. Targeted peptides included those released from *DIGESTIF*, and 4 signature peptides from 3 liver biomarkers (L-FABP, ALAT1, ADH1) (see Table 2). The upper chromatogram was obtained when trypsin was used alone for proteolysis (Test 1). The chromatogram for samples digested with trypsin/Lys-C mix in the presence of urea (Test 9) is presented in the lower panel. The y-axis scale is identical in both chromatograms to facilitate identification of differences in signal intensity. The peptides best suited for identification of differences in digestion efficiency are indicated by dotted lines between the two spectra.

Figure 7. Detection of liver biomarkers in plasma using *DIGESTIF* standard and LC-SRM.

A plasma sample from a patient with drug-induced liver injury was spiked with *DIGESTIF* standard. The plasma sample was digested using seven different proteolysis protocols (Tests 1, 7, 8, 9, 10, 12 and 13, see Table 1) and analyzed by LC-SRM, as described in the methods section. Label-free LC-SRM data were processed and SRM signal intensity (sum of peak area values obtained for 3 SRM transitions) was determined for each peptide in each digestion condition.

Figure 8. Multi-site evaluation of serum digestion protocols using *DIGESTIF* standard.

Serum sample preparation, spiking with *DIGESTIF*, sample digestion and LC-SRM analysis was performed in 3 different laboratories, as described in the experimental section. At each site, and for the 5 different proteolysis protocols (Table 1), 4 peptides (iRT peptides i and g, missed-cleaved iRT peptide g and PBP2x core peptide) were quantified by LC-SRM by comparison to unlabeled (AQUA) peptide standards. Each bar corresponds to the mean ± standard deviation calculated based on 3 SRM transitions.

Table 1. Protocols used for proteolysis

Test number	Enzyme	Enzyme/protein ratio (w/w)	Additive	Incubation		Remarks
				Temperature	Time	
Test 1	Trypsin	1/20	No	37°C	5 h	
Test 2	Trypsin	1/100	No	37°C	5 h	
Test 3	Trypsin	1/1000	No	37°C	5 h	
Test 4	Trypsin	1/20	No	30°C	5 h	
Test 5	Trypsin	1/20	No	23°C	5 h	
Test 6	Trypsin	1/20	No	37°C	5 h	Digestion was performed in sodium citrate (pH 2.2)
Test 7	Trypsin	1/20	No	37°C	24 h	
Test 8	Trypsin/Lys-C	1/20	No	37°C	5 h	
Test 9	Trypsin/Lys-C	1/20	Urea	37°C	5 h	Sample was incubated for 1 h 30 in 4 M urea. The urea concentration was reduced to 1 M by dilution, and digestion was allowed to proceed for 3 h 30.
Test 10	Trypsin	1/20	TCEP and iodoacetamide	37°C	5 h	Sample was reduced (TCEP 2 mM, 30 min) and alkylated (iodoacetamide 10 mM, 30 min in the dark) before trypsin digestion
Test 11	Trypsin/Lys-C	1/20	Urea	37°C	24 h	Sample was incubated for 3 h in 4 M urea. The urea concentration was reduced to 1 M by dilution, and digestion was allowed to proceed for 21 h

Table 1 (continued).

Test number	Enzyme	Enzyme/protein ratio (w/w)	Additive	Incubation		Remarks
				Temperature	Time	
Test 12	Trypsin	1/20	Rapigest 0.1%	37°C	5 h	
Test 13	Trypsin	1/20	Rapigest 0.1%	37°C	5 h	Sample was heated to 100 °C for 5 min before adding trypsin
Test 14	Trypsin	1/20	TCEP and iodoacetamide	37°C	5 h	FASP protocol was applied with reduction (TCEP 2 mM, 30 min) and alkylation (iodoacetamide 10 mM, 30 min in the dark). Trypsin digestion was allowed to proceed for 5 h.
Test 15	Trypsin/Lys-C	1/20	Urea	37°C	5 h	FASP protocol was applied without reduction and alkylation treatment. Digestion was performed 1 h 30 in 4 M urea. Then, urea concentration was reduced to 1 M and hydrolysis was allowed to proceed for 3 h 30.
Test 16	Trypsin/Lys-C	1/20	TCEP, iodoacetamide and urea	37°C	5 h	FASP protocol was applied with reduction (TCEP 2 mM, 30 min) and alkylation (iodoacetamide 10 mM, 30 min in the dark). Digestion was performed 1 h 30 in 4 M urea. Then, urea concentration was reduced to 1 M and hydrolysis was allowed to proceed for 3 h 30.

All digestion reactions were performed in 40 mM ammonium bicarbonate buffer (except for Test 6)

Table 2. Peptides investigated and their SRM transitions

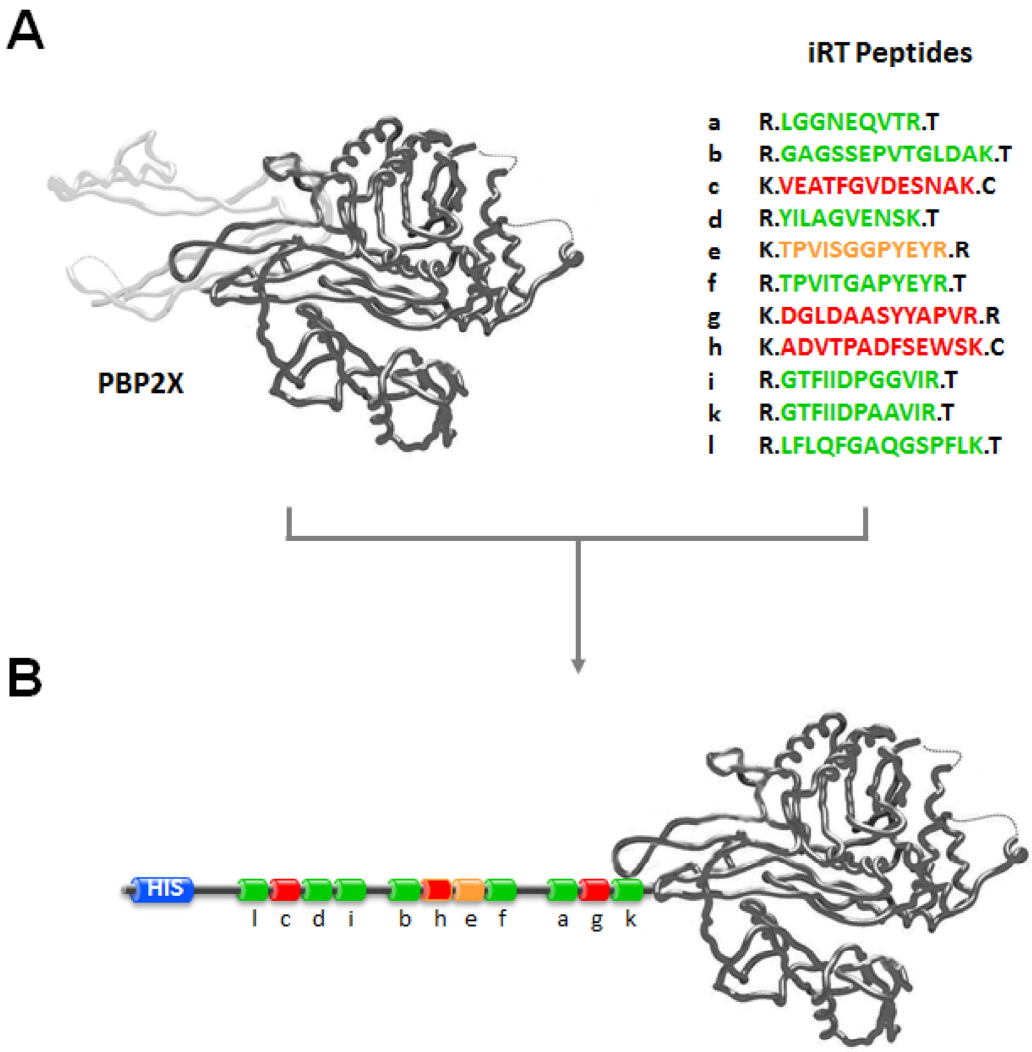
Peptide sequence	Peptide features	SRM transitions	
		Precursor Q1 m/z	Fragment Q3 m/z
LGGNEQVT**R**	iRT peptide a	492.3	870.4 (y8)
		492.3	813.4 (y7)
		492.3	642.3 (y5)
GAGSSEPVTGLDA**K**	iRT peptide b	648.8	937.5 (y9)
		648.8	808.5 (y8)
		648.8	612.3 (y6)
VEATFGVDESNA**K**	iRT peptide c	687.8	1146.5 (y11)
		687.8	974.5 (y9)
		687.8	827.4 (y8)
YILAGVENS**K**	iRT peptide d	551.3	825.5 (y8)
		551.3	712.4 (y7)
		551.3	641.3 (y6)
TPVISGGPYEY**R**	iRT peptide e	674.8	1051.5 (y9)
		674.8	938.4 (y8)
		674.8	851.4 (y7)
TPVITGAPYEY**R**	iRT peptide f	688.9	966.5 (y8)
		688.9	865.4 (y7)
		688.9	737.3 (y5)
DGLDAASYYAPV**R**	iRT peptide g	704.3	865.4 (y7)
		704.3	615.3 (y5)
		704.3	452.3 (y4)
DGLDAASYYAPVR	*iRT peptide g*	699.3	855.4 (y7)
		699.3	605.3 (y5)
		699.3	442.3 (y4)
TACox3KDGLDAASYYAPV**R**	Missed-cleaved iRT peptide g	623.0	778.4 (y6)
		623.0	615.4 (y5)
		623.0	452.3 (y4)
TACox^{3}KDGLDAASYYAPVR	*Missed-cleaved iRT peptide g*	617.0	768.4 (y6)
		617.0	605.4 (y5)
		617.0	442.3 (y4)
GTFIIDPGGVI**R**	iRT peptide i	627.9	836.5 (y8)
		627.9	723.4 (y7)
		627.9	608.4 (y6)
GTFIIDPGGVIR	*iRT peptide i*	622.9	826.5 (y8)
		622.9	713.4 (y7)
		622.9	598.4 (y6)
GTFIIDPAAVI**R**	iRT peptide k	641.9	864.5 (y8)
		641.9	751.4 (y7)
		641.9	636.4 (y6)

Labeled peptides from the DIGESTIF are mentioned with the C-terminal, isotope-labeled amino acid, in bold. Italicized peptides were used for absolute quantification (light unlabeled peptide standards). Cysteine trioxidation (ox^3) modification state is indicated.

Table 2 (continued).

Peptide sequence	Peptide features	SRM transitions	
		Precursor Q1 m/z	Fragment Q3 m/z
LFLQFGAQGSPFL**K**	iRT peptide l	780.9	1059.6 (y10)
		780.9	912.5 (y9)
		780.9	656.4 (y6)
NLVQPIVVGTGT**K**	PBP2x core peptide	667.4	1106.7 (y11)
		667.4	1007.6 (y10)
		667.4	879.5 (y9)
NLVQPIVVGTGTK	*PBP2x core peptide*	663.4	1098.7 (y11)
		663.4	999.8 (y10)
		663.4	871.5 (y9)
AIGLPEELIQK	L-FABP signature peptide	605.9	856.5 (y7)
		605.9	630.4 (y5)
		605.9	428.7 (y7)
TVVQLEGDNK	L-FABP signature peptide	551.8	902.5 (y8)
		551.8	675.3 (y6)
		551.8	562.2 (y5)
LLVAGEGHTR	ALAT1 signature peptide	526.8	727.3 (y7)
		526.8	656.3 (y6)
		526.8	413.7 (y8)
FSLDALITHVLPFEK	ADH1 signature peptide	577.3	791.9 (y14)
		577.3	748.4 (y13)
		577.3	691.9 (y12)

Figure 1



C

MGSSHHHHHHSSGLVPRGSRRASVHMGSGAKRGTIYDRNGVPIAEDATSGGARLFLQFGAQGSPFLKT
AHWKVEATFGVDESNAKCARYILAGVENSKTAARGTFIIDPGGVIRTAPNRSYPNGQFASSFIGGARGA
GSSEPVTGLDAKTAHWKADVTPADFSEWSKCHWKTPVISGGPYEYRRPARTPVITGAPYEYRTAGMES
SLNSILAGGGGARLGGNEQVTRTACKDGLDAASYYAPVRRPARGTFIIDPAAVIRTAELDGKDVYTTISSPL
QSFMETQMDAFQEKVKGKYMTATLVSAKTGEILATTQRPTFDADTKEGITEDFVWRDILYQSNYEPGSTM
KVMMLAAAIDNNTFPGGEVFNSSELKIADATIRDWDVNEGLTGGRMMTFSQGFAHSSNVGMTLLEQK
MGDATWLDYLNRFKFGVPTRFGLTDEYAGQLPADNIVNIAQSSFGQGISVTQTQMIRAFTAIANDGVML
EPKFISAIYDPNDQTARKSQKEIVGNPVSKDAASLTRTNMVLVGTDPVYGTMYNHSTGKPTVTVPGQNVA
LKSGTAQIADEKNGGYLVGLTDYIFSAVSMSPAENPDFILYVTVQQPEHYSGIQLGEFANPILERASAMKDSL
NLQTTAKALEQVSQQSPYPMPSVKDISPGDLAEELRRNLVQPIVVGTGTKIKNSSAEEGKNLAPNQQVLILS
DKAEEVPDMYGWTKETAETLAKWLNIELEFQGSGSTVQKQDVRANTAIKDIKKITLTLGD

Figure 2

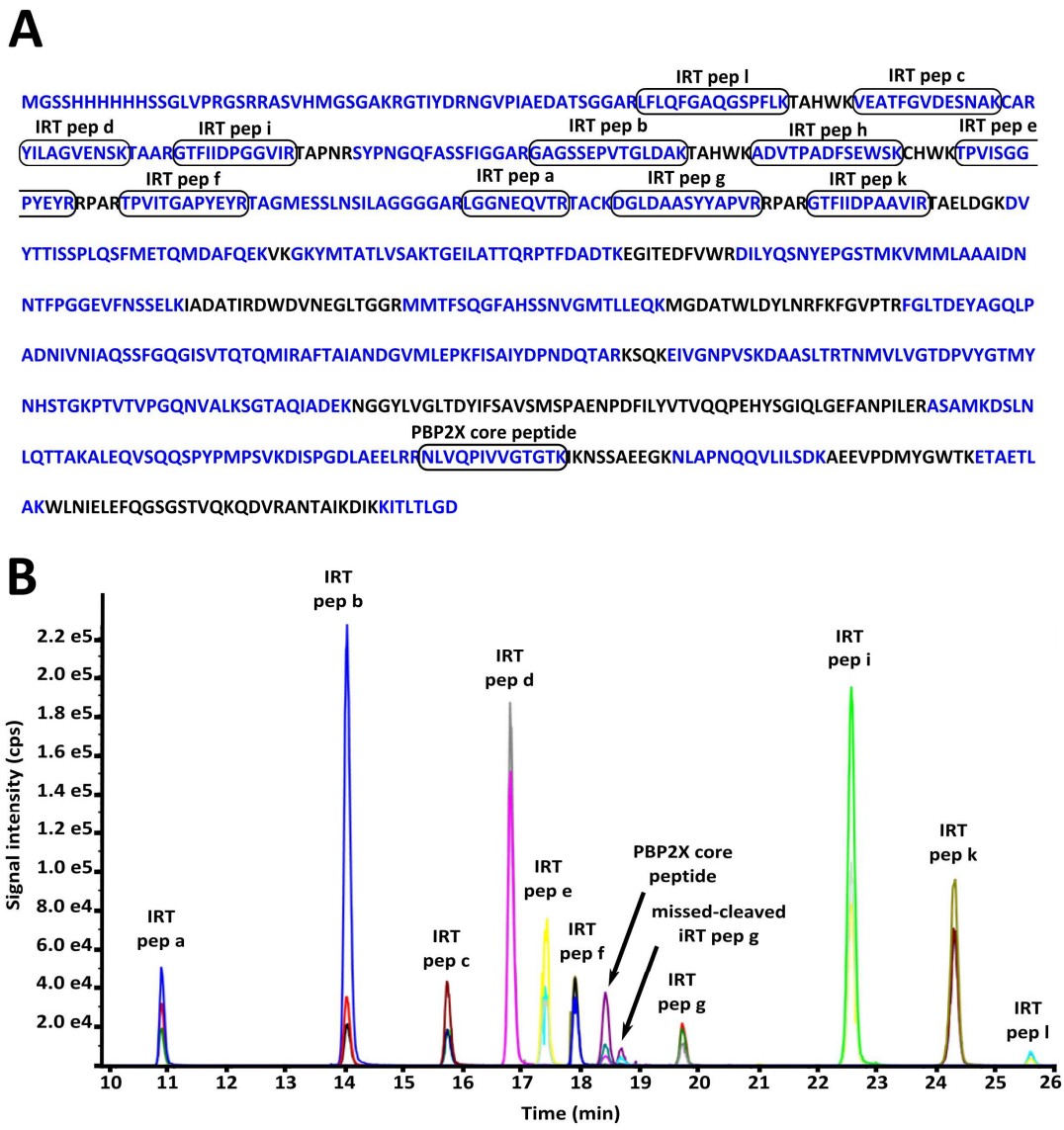

A
IRT pep l
IRT pep c
MGSSHHHHHHSSGLVPRGSRRASVHMGSGAKRGTIYDRNGVPIAEDATSGGARLFLQFGAQGSPFLKTAHWKVEATFGVDESNAKCAR
IRT pep d
IRT pep i
IRT pep b
IRT pep h
IRT pep e
YILAGVENSKTAARGTFIIDPGGVIRTAPNRSYPNGQFASSFIGGARGAGSSEPVTGLDAKTAHWKADVTPADFSEWSKCHWKTPVISGG
IRT pep f
IRT pep a
IRT pep g
IRT pep k
PYEYRRPARTPVITGAPYEYRTAGMESSLNSILAGGGGARLGGNEQVTRTACKDGLDAASYYAPVRRPARGTFIIDPAAVIRTAELDGKDV
YTTISSPLQSFMETQMDAFQEKVKGKYMTATLVSAKTGEILATTQRPTFDADTKEGITEDFVWRDILYQSNYEPGSTMKVMMLAAAIDN
NTFPGGEVFNSSELKIADATIRDWDVNEGLTGGRMMTFSQGFAHSSNVGMTLLEQKMGDATWLDYLNRFKFGVPTRFGLTDEYAGQLP
ADNIVNIAQSSFGQGISVTQTQMIRAFTAIANDGVMLEPKFISAIYDPNDQTARKSQKEIVGNPVSKDAASLTRTNMVLVGTDPVYGTMY
NHSTGKPTVTVPGQNVALKSGTAQIADEKNGGYLVGLTDYIFSAVSMSPAENPDFILYVTVQQPEHYSGIQLGEFANPILERASAMKDSLN
PBP2X core peptide
LQTTAKALEQVSQQSPYPMPSVKDISPGDLAEELRRNLVQPIVVGTGTKIKNSSAEEGKNLAPNQQVLILSDKAEEVPDMYGWTKETAETL
AKWLNIELEFQGSGSTVQKQDVRANTAIKDIKKITLTLGD
B
IRT pep b
IRT pep d
IRT pep i
IRT pep k
IRT pep a
IRT pep c
IRT pep e
IRT pep f
PBP2X core peptide
missed-cleaved iRT pep g
IRT pep g
IRT pep l
Signal intensity (cps)
2.2 e5
2.0 e5
1.8 e5
1.6 e5
1.4 e5
1.2 e5
1.0 e5
8.0 e4
6.0 e4
4.0 e4
2.0 e4
10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
Time (min)

Figure 3

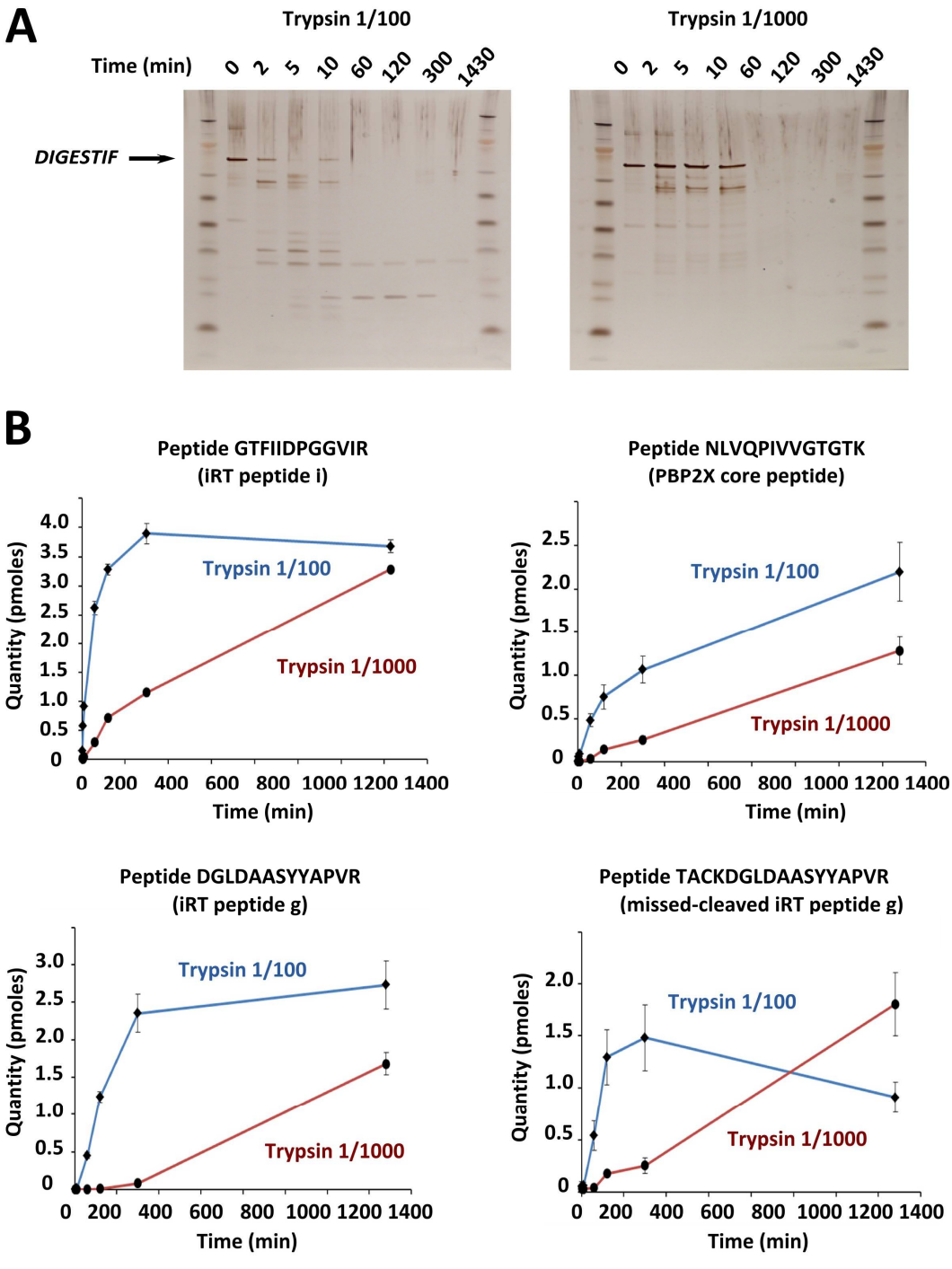
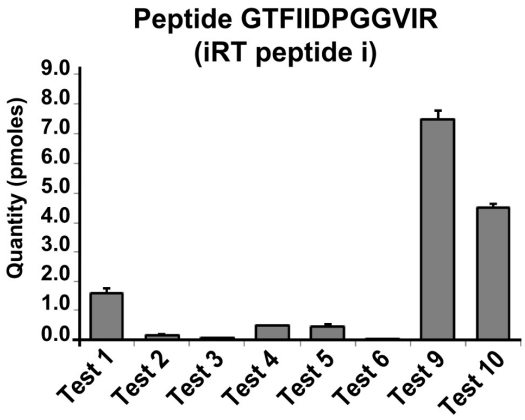
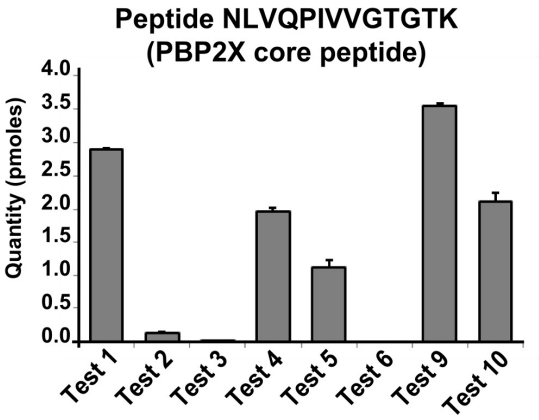
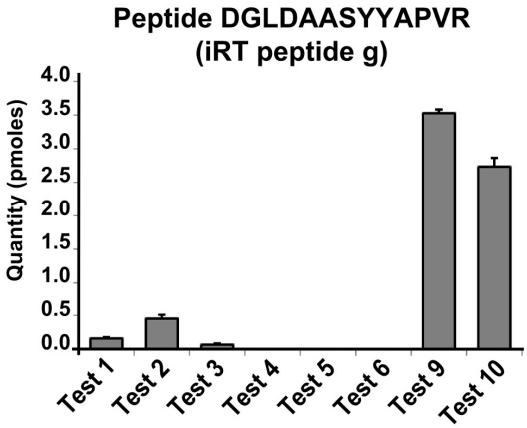

A
Trypsin 1/100
Time (min)
0 2 5 10 60 120 300 1430
DIGESTIF
Trypsin 1/1000
0 2 5 10 60 120 300 1430
B
Peptide GTFIIDPGGVIR
(iRT peptide i)
Quantity (pmoles)
Trypsin 1/100
Trypsin 1/1000
Time (min)
Peptide NLVQPIVVGTGTK
(PBP2X core peptide)
Quantity (pmoles)
Trypsin 1/100
Trypsin 1/1000
Time (min)
Peptide DGLDAASYYAPVR
(iRT peptide g)
Quantity (pmoles)
Trypsin 1/100
Trypsin 1/1000
Time (min)
Peptide TACKDGLDAASYYAPVR
(missed-cleaved iRT peptide g)
Quantity (pmoles)
Trypsin 1/100
Trypsin 1/1000
Time (min)

Figure 4

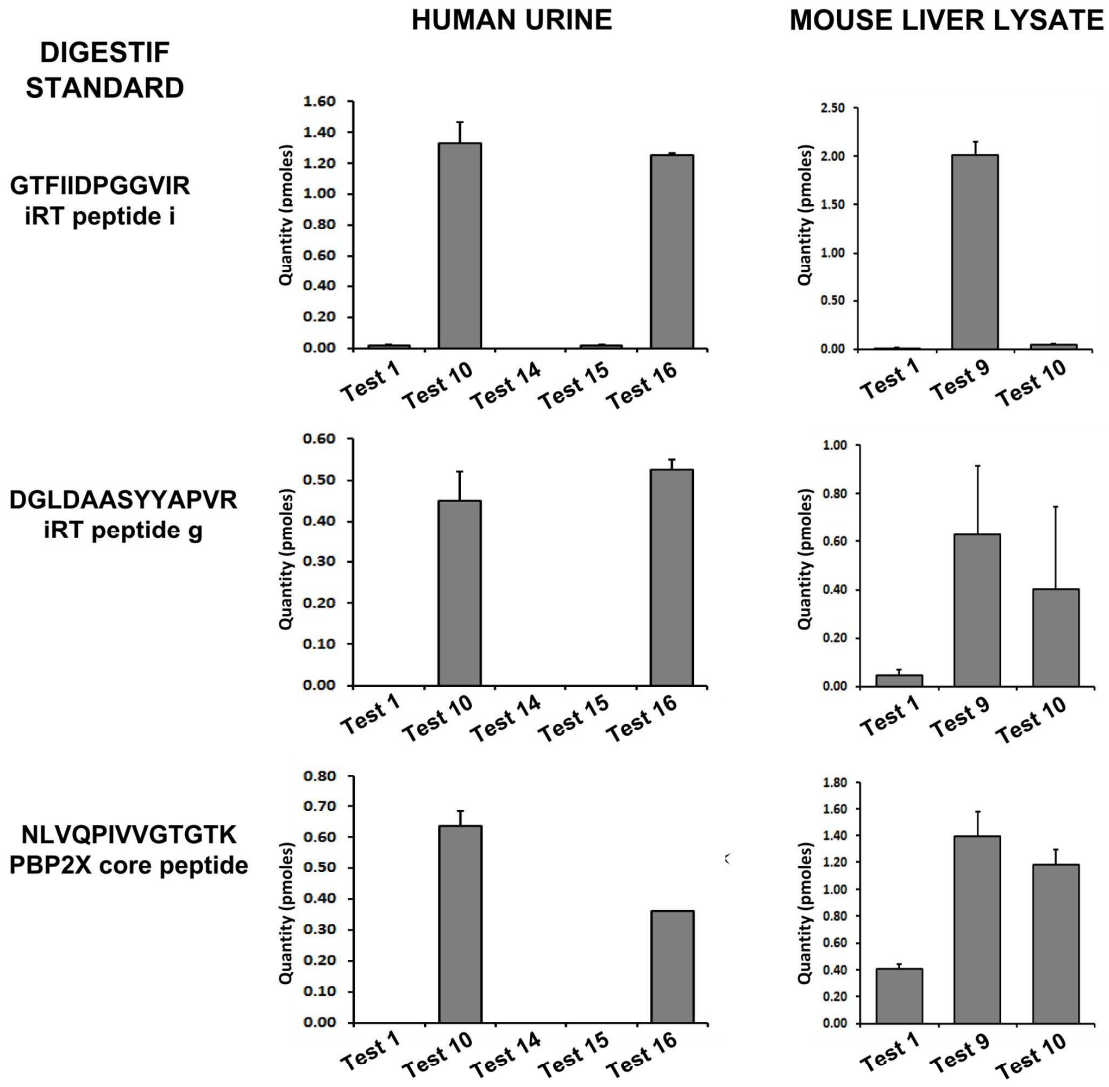






TEST	DIGESTION CONDITIONS
1	Trypsin 1/20, 37°C, 5h
2	Trypsin 1/100, 37°C, 5h
3	Trypsin 1/1000, 37°C, 5h
4	Trypsin 1/20, 30°C, 5h
5	Trypsin 1/20, 23°C, 5h
6	Trypsin 1/20, 37°C, 5h, pH 2.2
9	Trypsin/Lys-C 1/20, 37°C, 1h30 in 4M urea + 3h30 in 1M urea
10	Trypsin 1/20, 37°C, 5h after reduction/alkylation

Figure 5



TEST	DIGESTION CONDITIONS
1	Trypsin 1/20, 37°C, 5h
9	Trypsin/Lys-C 1/20, 37°C, 1h30 in 4M urea + 3h30 in 1M
10	Trypsin 1/20, 37°C, 5h after reduction/alkylation
14	Trypsin 1/20, 37°C, 5h, reduction/alkylation, FASP with urea
15	Trypsin/Lys-C, 37°C, 5h, FASP with urea
16	Trypsin/Lys-C, 37°C, 5h, reduction/alkylation, FASP with urea

Figure 6

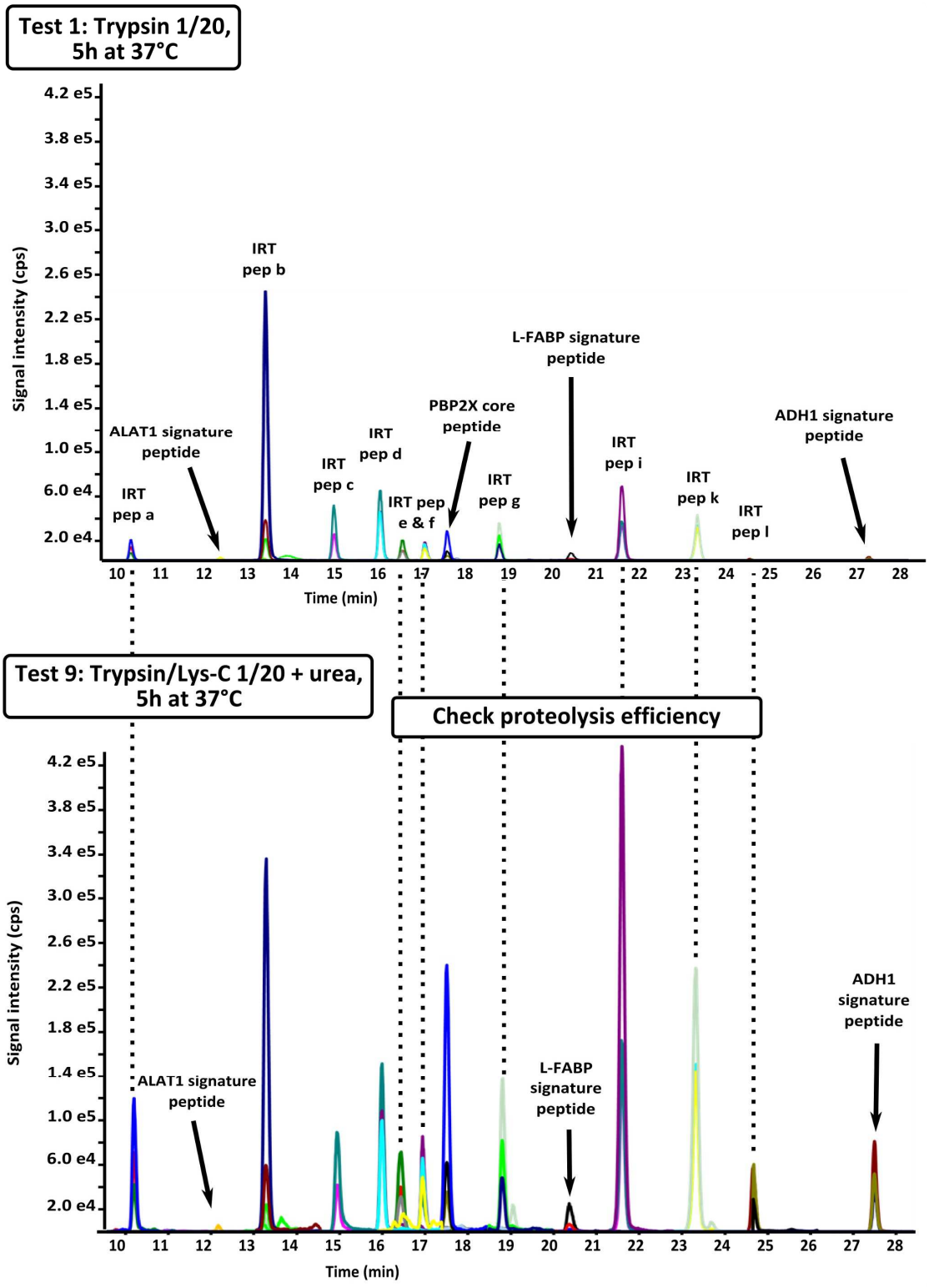

Test 1: Trypsin 1/20, 5h at 37°C
Signal intensity (cps)
4.2 e5
3.8 e5
3.4 e5
3.0 e5
2.6 e5
2.2 e5
1.8 e5
1.4 e5
1.0 e5
6.0 e4
2.0 e4
IRT pep a
ALAT1 signature peptide
IRT pep b
IRT pep c
IRT pep d
IRT pep e & f
PBP2X core peptide
IRT pep g
L-FABP signature peptide
IRT pep i
IRT pep k
IRT pep l
ADH1 signature peptide
10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
Time (min)
Test 9: Trypsin/Lys-C 1/20 + urea, 5h at 37°C
Check proteolysis efficiency
Signal intensity (cps)
4.2 e5
3.8 e5
3.4 e5
3.0 e5
2.6 e5
2.2 e5
1.8 e5
1.4 e5
1.0 e5
6.0 e4
2.0 e4
ALAT1 signature peptide
L-FABP signature peptide
ADH1 signature peptide
10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28
Time (min)

Figure 7

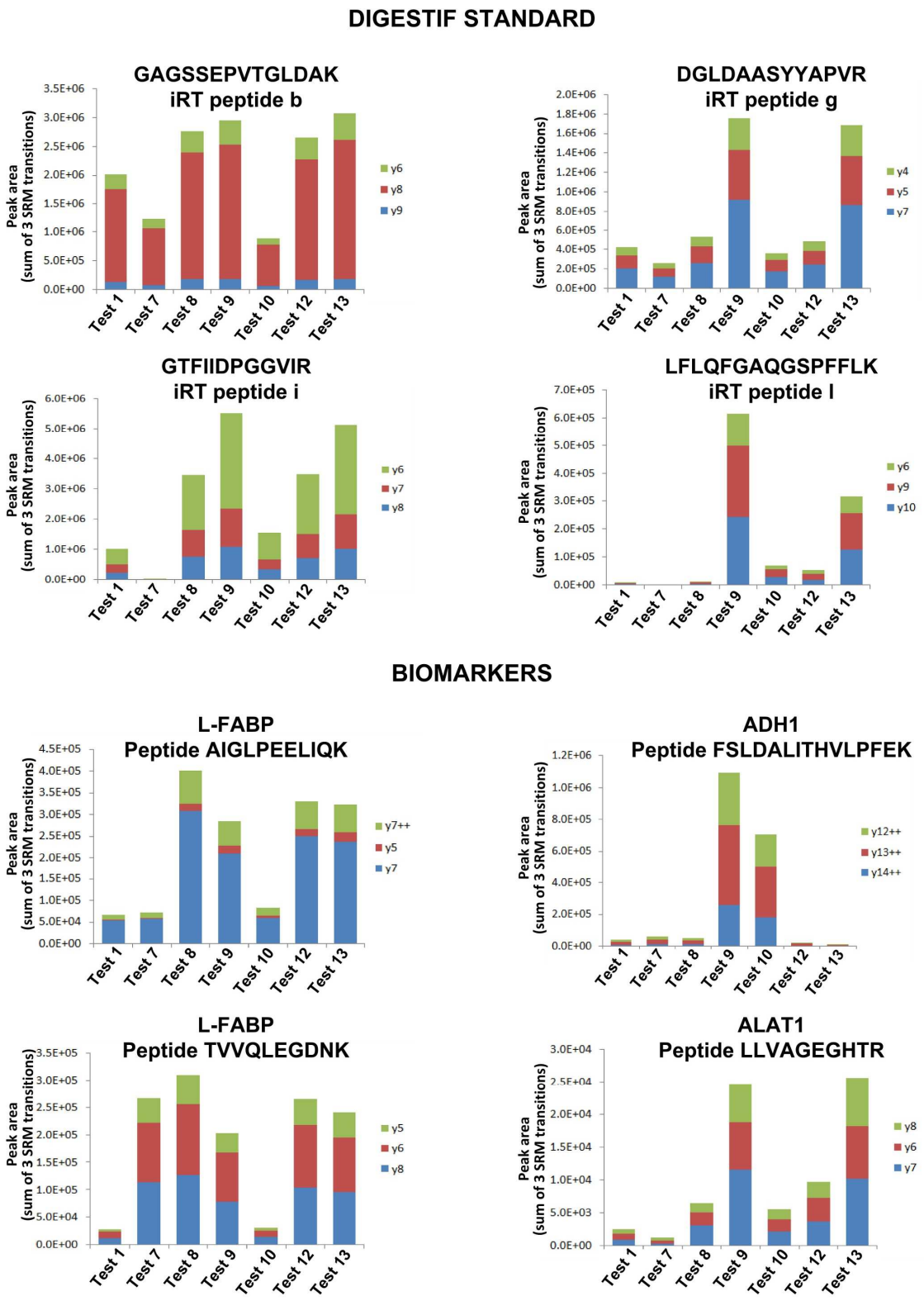
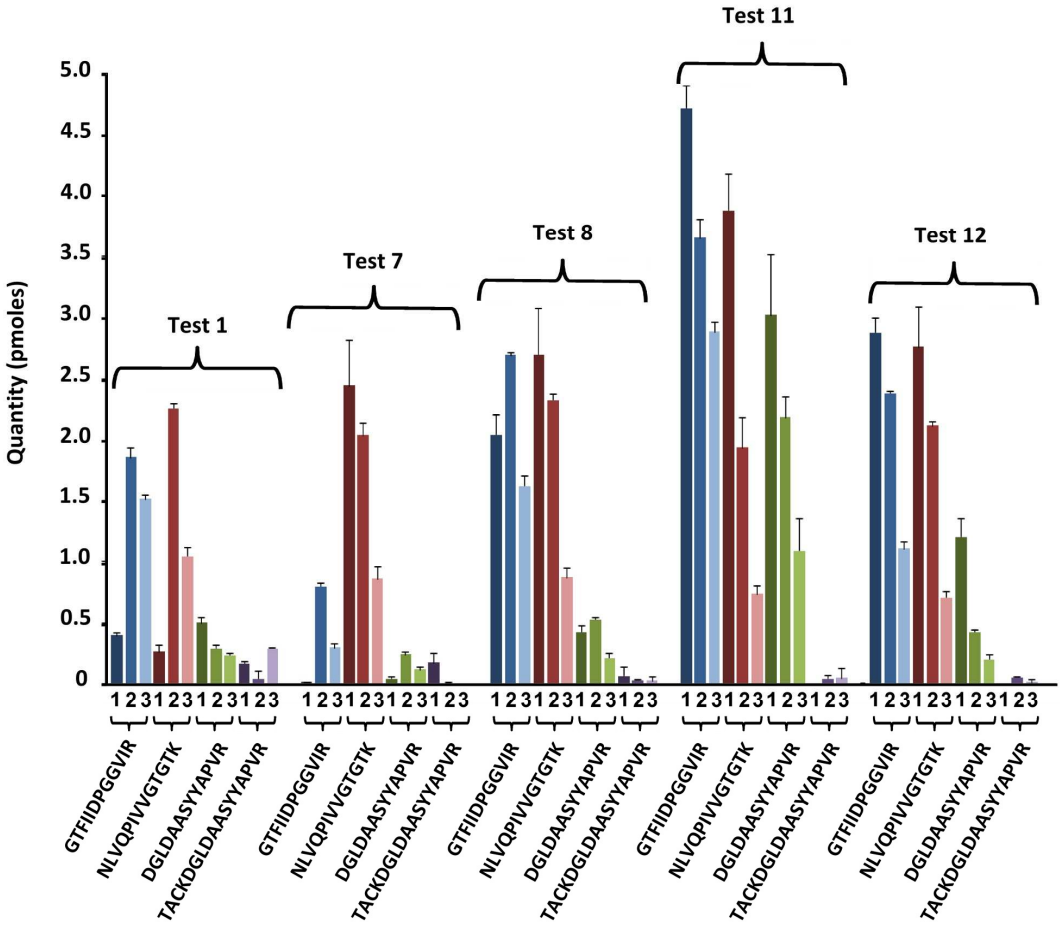

DIGESTIF STANDARD
GAGSSEPVTGLDAK
iRT peptide b
Peak area
(sum of 3 SRM transitions)
y6
y8
y9
DGLDAASYYAPVR
iRT peptide g
y4
y5
y7
GTFIIDPGGVIR
iRT peptide i
y6
y7
y8
LFLQFGAQGSPFFLK
iRT peptide l
y6
y9
y10
Test 1
Test 7
Test 8
Test 9
Test 10
Test 12
Test 13
BIOMARKERS
L-FABP
Peptide AIGLPEELIQK
y7++
y5
y7
ADH1
Peptide FSLDALITHVLPFEK
y12++
y13++
y14++
L-FABP
Peptide TVVQLEGDNK
y5
y6
y8
ALAT1
Peptide LLVAGEGHTR
y8
y6
y7

Figure 8


Test 11
Test 1
Test 7
Test 8
Test 12
Quantity (pmoles)
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0
1 2 3 1 2 3 1 2 3 1 2 3
GTFIIDPGGVIR
NLVQPIVVGTGTK
DGLDAASYYAPVR
TACKDGLDAASYYAPVR

SUPPORTING INFORMATION

Supplementary Figures S1 and S2.

S1: *DIGESTIF* quality controls.

S2: Classification of iRT peptides

Supplementary Table ST1. Label-free analysis of peptide release during *DIGESTIF* proteolysis.

Supplementary Table ST2. Analysis of *DIGESTIF* peptide release in human serum (first sheet), human urine (second sheet) and mouse liver lysate (third sheet).

Supplementary Table ST3. *DIGESTIF* and liver injury biomarker proteolysis in serum: label-free analysis of peptide release.

Supplementary Table ST4. Complete data for multi-site assessment of *DIGESTIF* proteolysis in serum.

This material is available free of charge via the Internet at http://pubs.acs.org